



ARL-TR-8353 • MAY 2018



Intercomparison of 7 Planetary Boundary-Layer/Surface-Layer Physics Schemes over Complex Terrain for Battlefield Situational Awareness

by Richard S Penc, Jeffrey A Smith, John W Raby, Robert E Dumais Jr, Brian P Reen, and Leelinda P Dawson

Approved for public release; distribution is unlimited.

NOTICES

Disclaimers

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Citation of manufacturer's or trade names does not constitute an official endorsement or approval of the use thereof.

Destroy this report when it is no longer needed. Do not return it to the originator.



Intercomparison of 7 Planetary Boundary-Layer/Surface-Layer Physics Schemes over Complex Terrain for Battlefield Situational Awareness

by Richard S Penc, Jeffrey A Smith, John W Raby, Robert E Dumais Jr, Brian P Reen, and Leelinda P Dawson
Computational and Information Sciences Directorate, ARL

REPORT DOCUMENTATION PAGE				Form Approved OMB No. 0704-0188	
<p>Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.</p> <p>PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.</p>					
1. REPORT DATE (DD-MM-YYYY) April 2018		2. REPORT TYPE Technical Report		3. DATES COVERED (From - To) 1 January 2015–26 February 2018	
4. TITLE AND SUBTITLE Intercomparison of 7 Planetary Boundary-Layer/Surface-Layer Physics Schemes over Complex Terrain for Battlefield Situational Awareness				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) Richard S Penc, Jeffrey A Smith, John W Raby, Robert E Dumais Jr, Brian P Reen, and Leelinda P Dawson				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) US Army Research Laboratory ATTN: RDRL-CIE-M Aberdeen Proving Ground, MD 21005-5067				8. PERFORMING ORGANIZATION REPORT NUMBER ARL-TR-8353	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT This study considers the performance of 7 of the Weather Research and Forecast model boundary-layer (BL) parameterization schemes in a complex terrain environment. These BL/surface-layer options were compared and represent local, nonlocal, and hybrid approaches. Based on surface parameter root-mean-square error and bias measures, there is no clearly superior scheme; however, the BouLac, Quasi-normal Scale Elimination, and Yonsei University (YSU) schemes performed best. The surface parameters, planetary BL structure, and vertical profiles are important for US Army Research Laboratory applications, including in the formulation of some weather impact decision aids. Both the YSU and Mellor-Yamada-Janjic schemes, used in the common Weather Running Estimate–Nowcast model configurations and in operational configurations, appear to perform reasonably well in this environment. While the scale-aware scheme addresses the scale dependency problem, we have found no clear improvement when using this scheme with a 1-km inner mesh in this study. There is little variation between the schemes in the daytime BL, but at nighttime the schemes differ notably. The model's greatest difficulty is in simulating the transition between daytime and nighttime BLs. There is insufficient statistical spread in the members tested to consider a BL-physics-only-based ensemble to quantify model uncertainty.					
15. SUBJECT TERMS numerical weather prediction, Weather Research and Forecast (WRF) Model, planetary boundary layer, parameterizations, ensemble, complex terrain, nowcast					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT UU	18. NUMBER OF PAGES 112	19a. NAME OF RESPONSIBLE PERSON Richard Penc
a. REPORT Unclassified	b. ABSTRACT Unclassified	c. THIS PAGE Unclassified			19b. TELEPHONE NUMBER (Include area code) 575-678-0105

Contents

List of Figures	v
List of Tables	viii
Preface	x
Acknowledgments	xi
Executive Summary	xiii
1. Introduction	1
1.1 Research Objectives	2
1.2 Common Model Configurations: Subgrid-Scale Physics	4
1.3 Structure of Report	7
2. Planetary Boundary Layer	7
2.1 Boundary Layer Development and Structure	8
2.2 Features within the Convective Boundary Layer	10
2.3 Sources of Model Error	12
3. WRF Boundary-Layer/Surface-Layer Physics Packages	14
3.1 WRF v.3.8.1 Boundary-Layer/Surface-Layer Schemes	14
3.1.1 Yonsei University Scheme (YSU)	16
3.1.2 Mellor-Yamada-Janjic Scheme (MYJ)	17
3.1.3 National Center for Environmental Prediction (NCEP) Global Forecast System Scheme	17
3.1.4 Quasi-normal Scale Elimination (QNSE) Scheme	17
3.1.5 Mellor-Yamada Nakanishi Niino (MYNN) Levels 2.5 and 3 Schemes	18
3.1.6 Asymmetric Convection Model 2 Scheme (ACM2)	18
3.1.7 Bougeault–Lacarrère Scheme (BouLac)	19
3.1.8 University of Washington (UW) TKE Boundary-Layer Scheme	19

3.1.9	Total Energy Mass Flux (TEMF) Surface Layer Scheme	19
3.1.10	Shin-Hong Scale-Aware Scheme	20
3.1.11	Grenier-Bretherton-McCaa Scheme	20
3.1.12	Medium Range Forecast Scheme	20
3.2	Other PBL Scheme Intercomparison Studies	21
3.3	Direction	23
4.	Domain and Model Configuration	24
5.	Results	31
5.1	Model Bias and RMSE	32
5.1.1	Model Bias	33
5.1.2	Model RMSE	41
5.1.3	Summary Statistics	47
5.2	Talagrand Diagrams	52
5.3	Horizontal Variation in Surface (2m) Temperature	56
5.4	PBL Depth	64
5.5	Computation Time	72
5.6	Analysis Summary	73
6.	Summary and Conclusions	75
7.	Further Considerations	78
8.	References	82
	List of Symbols, Abbreviations, and Acronyms	93
	Distribution List	97

List of Figures

Fig. 1	Horizontal grid-scale dependence of convection and turbulence parameterizations (Deng and Stauffer 2006). Cross-hatched areas indicate where the underlying assumptions of the physical parameterizations are not strictly valid, and there is currently no robust, satisfactory solution in these no-man's land scale ranges: convection (4–10 km Δx) and turbulence ($\Delta x \sim$ depth of mixed layer). 6
Fig. 2	Diurnal evolution and cycle in PBL structure over land and under quiescent (fair weather) conditions 8
Fig. 3	BL plume structures. The vertical dimension is determined by the depth of the BL, and is typically equal to or larger than the horizontal dimension of the plumes. Smaller-scale turbulent motions exist within the cloud structure..... 10
Fig. 4	Structure of horizontal BL rolls, showing the circulations that develop and associated sinusoidal behavior of the vertical motion field. The wavelength of these rolls is approximately 2–5 km. Cloud streets form in sufficiently moist BLs in the upward circulation between rolls.. ... 11
Fig. 5	Steps in the running of WRF-ARW at ARL, showing the ingest of meteorological and coarse grid data, preprocessing, and production of a WRF forecast preceded by a preforecast with observational nudging. In this study, the preforecast was 6 h and the forecast 18 h..... 27
Fig. 6	Map of the model domain for these experiments. The configuration consists of a triple nested domain (9-3-1 km) centered approximately over San Diego, CA. It is the same domain used for the ARL GIS studies. 28
Fig. 7	Analysis (500 hPa) for 1200Z 09 February 2012 (left) and 0000Z 10 February 2012 (right)..... 30
Fig. 8	Surface weather maps for 1200Z 09 February 2012 (left) and 0000Z 10 February 2012 (right)..... 30
Fig. 9	Time series plot of the surface (2m) winds at KNZY; latitude 32.69°N, longitude 117.21°W, and elevation 23 ft 31
Fig. 10	Model bias for surface temperature (Kelvin [K]) at 2m above ground level (AGL) for the 7 PBL schemes tested. The data assimilation was applied from 1200 to 1800 UTC, and thus 0000 UTC is the beginning of the nowcast. 34
Fig. 11	Model bias for DPT (K) at 2m AGL for the 7 PBL schemes tested. The data assimilation was applied from 1200 to 1800 UTC, and thus 0000 UTC is the beginning of the nowcast..... 35
Fig. 12	Model bias for RH (%) at 2m AGL for the 7 PBL schemes tested. The data assimilation was applied from 1200 to 1800 UTC, and thus 0000 UTC is the beginning of the nowcast..... 36

Fig. 13	Model bias for u component of wind (ms^{-1}) at 10m AGL for the 7 PBL schemes tested. The data assimilation was applied from 1200 to 1800 UTC, and thus 0000 UTC is the beginning of the nowcast.	38
Fig. 14	Model bias for wind u component (ms^{-1}) at 10m AGL for the 7 PBL schemes tested. The data assimilation was applied from 1200 to 1800 UTC, and thus 0000 UTC is the beginning of the nowcast.	39
Fig. 15	Model bias for wind velocity (ms^{-1}) at 10m AGL for the 7 PBL schemes tested. The data assimilation was applied from 1200 to 1800 UTC, and thus 0000 UTC is the beginning of the nowcast.	40
Fig. 16	Model RMSE for surface temperature (K) at 2m AGL for the 7 PBL schemes tested. The data assimilation was applied from 1200 to 1800 UTC, and thus 0000 UTC is the beginning of the nowcast.	42
Fig. 17	Model RMSE for DPT (K) at 2m AGL for the 7 PBL schemes tested. The data assimilation was applied from 1200 to 1800 UTC, and thus 0000 UTC is the beginning of the nowcast.....	43
Fig. 18	Model RMSE for RH (%) at 2m AGL for the 7 PBL schemes tested. The data assimilation was applied from 1200 to 1800 UTC, and thus 0000 UTC is the beginning of the nowcast.....	44
Fig. 19	Model RMSE for u component of wind (ms^{-1}) at 10m AGL for the 7 PBL schemes tested. The data assimilation was applied from 1200 to 1800 UTC, and thus 0000 UTC is the beginning of the nowcast.	45
Fig. 20	Model RMSE for v component of wind (ms^{-1}) at 10m AGL for the 7 PBL schemes tested. The data assimilation was applied from 1200 to 1800 UTC, and thus 0000 UTC is the beginning of the nowcast.	46
Fig. 21	Model RMSE for wind velocity (ms^{-1}) at 10m AGL for the 7 PBL schemes tested. The data assimilation was applied from 1200 to 1800 UTC, and thus 0000 UTC is the beginning of the nowcast.	47
Fig. 22	WRF D3 (1-km) ensemble rank histogram produced for 00Z 10 February 2012, surface (2m) temperature.....	53
Fig. 23	WRF D3 (1-km) ensemble rank histogram produced for 00Z 10 February 2012, surface (2m) DPT	54
Fig. 24	WRF D3 (1-km) ensemble rank histogram produced for 00Z 10 February 2012, 2m (surface) RH	54
Fig. 25	WRF D3 (1-km) ensemble rank histogram produced for 00Z 10 February 2012, 10m u component of wind.....	55
Fig. 26	WRF D3 (1-km) ensemble rank histogram produced for 00Z 10 February 2012, 10m v component of wind.....	55
Fig. 27	WRF D3 (1-km) ensemble rank histogram produced for 00Z 10 February 2012, 10m wind speed.....	56

Fig. 28	Terrain and location of the weather stations that contain persistent data for a related study over the innermost domain (1-km grid) created from a US Geological Survey digital elevation model with a resolution of 1/3 arc-second. A subset of these observations is used based on persistence through model integration.	57
Fig. 29	Ensemble mean surface (2m AGL) temperature (Kelvin) over the innermost model domain (D3, 1-km grid) at 0000 UTC 10 February 2012.....	58
Fig. 30	Ensemble maximum surface (2m AGL) temperature (Kelvin) over the innermost model domain (D3, 1-km grid) at 0000 UTC 10 February 2012.....	59
Fig. 31	Ensemble minimum surface (2m AGL) temperature (Kelvin) over the innermost model domain (D3, 1-km grid) at 0000 UTC 10 February 2012.....	60
Fig. 32	Ensemble range, surface (2m AGL) temperature (Kelvin) over the innermost model domain (D3, 1km grid) at 0000 UTC 10 February 2012.....	61
Fig. 33	Ensemble standard deviation of surface (2m AGL) temperature (Kelvin) over the innermost model domain (D3, 1-km grid) at 0000 UTC 10 February 2012	62
Fig. 34	Ensemble plus 1 σ of surface (2m) temperature (Kelvin) over the innermost model domain (D3, 1-km grid) at 0000 UTC 10 February 2012.....	63
Fig. 35	Ensemble minus 1 standard deviation of surface (2m AGL) temperature (Kelvin) over the innermost model domain (D3, 1-km grid) at 0000 UTC 10 February 2012	63
Fig. 36	Atmospheric soundings at KNKX, San Diego (top) 0000 UTC 10 February 2012 and (bottom) 1200 UTC 10 February 2012	65
Fig. 37	Temporal evolution of PBL depth for the 7 PBL schemes tested over the center point of innermost (D3, 1-km) domain	66
Fig. 38	PBL depth (meters) valid at 0000 UTC 10 February 2012 for the MYNN PBL option.....	68
Fig. 39	PBL depth (meters) valid at 0000 UTC 10 February 2012 for the SH PBL option	69
Fig. 40	PBL depth (meters) valid at 0000 UTC 10 February 2012 for the MYJ PBL option	69
Fig. 41	PBL depth (meters) valid at 0000 UTC 10 February 2012 for the YSU PBL option	70
Fig. 42	PBL depth (meters) valid at 0000 UTC 10 February 2012 for the BouLac PBL option	70
Fig. 43	PBL depth (meters) valid at 0000 UTC 10 February 2012 for the QNSE PBL option.....	71

Fig. 44	PBL depth (meters) valid at 0000 UTC 10 February 2012 for the ACM2 PBL option	72
Fig. 45	Execution time (hours) for a 24-h simulation using the triple nest configuration and each of the 7 tested PBL/SL scheme combinations	73
Fig. 46	Sample of WRF-ARW output after assimilation into AIR. Based on user-specified thresholds, air routing algorithms provided an optimal flight path to minimize hazards. Environmental parameters were provided by WRF-ARW output from this study.	81

List of Tables

Table 1	Common physics scheme/option configurations used by ARL, Air Force, and NSSL as of this writing	4
Table 2	Planetary BL and SL options available in WRF v3.8.1 including pairing information, prognostic and diagnostic variables available in each scheme, and whether it is a local, nonlocal, or hybrid scheme... 15	
Table 3	WRF-ARW common namelist options used for this study	24
Table 4	Initialization and nudging data switches (aside from coarse grid model data) used in the preprocessing package RUNWPSPLUS for WRF-ARW runs examined in this study	26
Table 5	Synoptic conditions for the case study days considered	29
Table 6	PBL/SL combinations tested and evaluated in this study	32
Table 7	Mean bias and RMSE for 1-km ensemble members (D3) (1 km) for 1200 UTC 9 February through 1800 UTC 9 February (data assimilation, 0–6 h)	48
Table 8	Mean bias and RMSE for 1-km ensemble members (D3) (1 km) for 1800 UTC 9 February through 0000 UTC 10 February (nowcast period, 6–12 h)	49
Table 9	Mean bias and RMSE for 1-km ensemble members (D3) (1 km) for 0000 UTC 10 February through 1200 UTC 10 February (extended forecast, 12–24 h)	49
Table 10	Mean bias and RMSE for 1-km ensemble members (D3) (1 km) for 1200 UTC 9 February through 1200 UTC 10 February (complete forecast, 6–24 h)	50
Table 11	Summary RMSE statistics (mean and standard deviation) and relative ranking of each of the PBL scheme configurations we tested for the nowcast, extended forecast, and total forecast periods. Overall ranking	

	for each scheme (COMP) is a composite of T, DPT, and wind speed rankings. Temperature and DPT units are in Kelvin and wind speed in ms^{-1}	51
Table 12	Temperature bias (2m AGL) and RMSE for 1- and 3-km ensemble members (D3/D2) (1 km/3 km) validation time: 0000 UTC on 10 February (12-h simulation time, and 6-h lead time)	52

Preface

This technical report explores planetary boundary-layer (BL) and surface-layer (SL) options for the Weather Research and Forecasting (WRF) model, which is the core of the Weather Running Estimate–Nowcast model (WRE-N), designed to provide forward-deployed locations’ weather and environmental data. This information is intended for integration into the common operational picture for decision making and planning purposes in forward-deployed locations. Previous work at the US Army Research Laboratory (ARL) involved the design of the system and looked at improvements to the verification process. Those results are published in the following ARL technical reports: Smith et al. (2015), Kirby et al. (2013), and Dumais et al. (2013).

The experiments described in this technical report are a subset of those used by Smith and Penc (2015a, 2015b, 2016) and Smith et al. (2017, 2018) in the design of experiments (DoE) research that explores a more systematic novel approach to the selection of parameterization schemes for nowcasting with WRF. In this report we examine the BL schemes only because they have the potential for providing for the greatest impact in the model physics uncertainty using the WRF model. The uncertainty associated with the other physics options (radiation, cumulus, and microphysics schemes) are explored in greater detail in the previously mentioned DoE research. The selection and performance of the BL and SL physics packages are explored in this report. This work is part of a larger effort to quantify the uncertainty in numerical weather prediction models, namely the WRF. WRE-N takes coarser model data from the operational weather centers (e.g., US Air Force 557th Weather Wing) and produces high-resolution local short-range forecasts to provide support for Army users, such as field deployed locations.

Acknowledgments

Richard Penc offers his thanks to Dr Steven Silberberg of the Aviation Weather Center, National Oceanic and Atmospheric Administration (NOAA), for his valuable suggestions and editing of the manuscript for structure and scientific content at various stages since its inception to completion. Thanks are also due to Mr Pieter Haines, whose tireless work supporting the computing facilities, particularly the local modeling server clusters, is greatly appreciated by the US Army Research Laboratory (ARL) Atmospheric Modeling Branch at White Sands Missile Range. He also thanks Dr Theresa Foley for the many lengthy and thoughtful discussions pertaining to Geographic Information Systems, specifically the potential sources of errors due to geolocation and coordinate transformations related to use of the National Center for Atmospheric Research Model Evaluation Tools analysis packages with our databases. Thanks are due to Jeff Johnson of ARL for providing the Air Impacts Routing figure and for invaluable discussions on interfacing Weather Running Estimate–Nowcast mode to decision tools.

AirDat LLC provided Tropospheric Airborne Meteorological Data Reporting observational data, which expanded the above-surface data available for data assimilation and verification. This study was made possible in part due to the data made available to NOAA by various providers for inclusion in the Meteorological Assimilation Data Ingest System (MADIS). The parallel Real-Time Mesoscale Analysis “use” and “reject” lists were provided by Steve Levine at the National Weather Service’s National Centers for Environmental Prediction’s Environmental Modeling Center and greatly facilitated making full use of the MADIS observational dataset.

Many thanks go to the ARL Technical Publishing Office for their consistently high standard of editing, technical review, and making things look nice.

INTENTIONALLY LEFT BLANK.

Executive Summary

The US Army has a demonstrated need for up-to-date, high temporal- and spatial-resolution meteorological data and short-term weather forecasts to support mission decision making and planning in field-deployed locations. As such, the Weather Running Estimate–Nowcast (WRE-N; Dumais et al. 2009) system was developed to address these needs. At the heart of WRE-N is the Weather Research and Forecast (WRF) model, which is a community open-source model developed primarily by the National Center for Atmospheric Research (Skamarock et al. 2008). The WRF model is capable of producing global and regional scale forecasts, as well as longer-term decadal climate simulations. Spatially, the model physics are capable of resolving large-scale synoptic weather systems, mesoscale phenomena such as severe storms, and sub-kilometer-scale phenomena down to the larger boundary-layer (BL) eddies through inclusion of a large-eddy simulation model. In addition, the model includes a wide array of physics packages that can be selected by the user, parameterizing subgrid-scale processes including solar (or shortwave, SW) and longwave (LW) radiation, cumulus (CU) and shallow cumulus (SCU), cloud microphysics (MP), planetary BL (PBL), surface layer (SL), and integrated land surface model options. The WRF stands out among other operational weather forecast models in its flexibility of choice in physical parameterizations.

As a community model, the WRF code and physics are not static. Aside from bug fixes and performance improvements, the modeling system is continually being updated with the latest physical parameterization schemes. The model is updated yearly, with a release 4 months later primarily to address bugs in the code. The current release of WRF (at time of this research) is 3.8, and the latest “bug-fixed” version is 3.8.1, released in August 2016. In the WRF v.3.8.1 release, there are 12 PBL, 8 SL, 6 LW, 7 SW, 21 cloud MP, 11 CU, and 3 SCU options selectable by the user. Along with tunable model parameters (e.g., diffusion) there are almost limitless combinations of schemes and parameterization schemes to choose from. Analysis of model output generated using these schemes is the subject of Design of Experiment (DoE) studies ongoing at the US Army Research Laboratory’s Atmospheric Modeling Branch (Smith and Penc 2015a,b, 2016) and Smith et al. (2017, 2018).

The present study focuses on the PBL/SL combinations because of the importance of simulating BL processes, which drive atmospheric motions near the surface where ground forces operate, where artillery meteorology impacts are great, and where land mobility is impacted. Of the 12 PBL schemes in WRF, 7 are examined for their overall performance in a complex terrain environment. The model domain

is in southern California, approximately centered near San Diego, California. It was chosen primarily because of the availability of high-density surface observations and diversity of terrain. A triple nest configuration (9–3–1 km) of WRF v3.8.1 was run for a single day in which conditions were quiescent in order to observe the evolution of the PBL in the absence of rapidly changing synoptic-scale weather influences. The present analysis is a subset of those used in the DoE study.

This study examines the PBL/SL performance in complex terrain, driven by the need for the WRE-N modeling system to be preconfigured to run optimally in nonpredisclosed forward-deployed locations for the purpose of battlefield situational awareness. The Mellor-Yamada-Nakanishi-Niino, Mellor-Yamada-Janjic, Yonsei University (YSU), Assymetric Convection Model 2, Bougeault-Lacarrère (BouLac), Quasi-normal Scale Elimination (QNSE), and Shin-Hong PBL/SL options were compared, representing local, nonlocal, and hybrid schemes. Postprocessing was performed using the Model Evaluation Tools package, which was designed to evaluate WRF forecasts. We constructed a 7-member physics-based ensemble for our analyses to examine the uncertainty due to PBL/SL physics. We calculated the root-mean-square error and bias for each of the schemes. Of the 7 schemes we analyzed, no single scheme stands out as universally optimal for use for our model domain and initial conditions. The standard deviation within the model ensemble members was small. While the performance of these schemes does not differ significantly for the daytime convective BL, greater variation is encountered during nighttime. The model appears to have the largest difficulty in simulating the BLs during the transition between daytime and nighttime.

Although the best performance is a function of meteorological variable (wind, temperature, and moisture) and time of day, overall the best performing schemes include the BouLac, YSU, and QNSE parameterizations. Analysis of our results from a single case day using Talagrand diagrams shows that there is insufficient spread in the members for the consideration of a PBL-scheme-only model ensemble. The near-surface meteorological parameters, PBL structure, and vertical profiles are important for our application and for the prediction of weather effects on Army operations using weather impacts decision aids. While the scale-aware scheme (Shin and Hong 2013) addresses the scale dependency problem, there was no notable significant improvement gained in using this scheme with the current 1-km inner-domain grid spacing.

1. Introduction

The US Army Research Laboratory (ARL) has been performing long-term research into application of the Advanced Research version of the Weather Research and Forecast (WRF-ARW) model (Skamarock et al. 2008) for battlefield short-range forecasting in a field-deployed location. The purpose of these forecasts is to provide weather support for mission planning and execution and to augment coarser resolution models for specific applications required by field deployed units. Currently, the resolution of the operational weather modeling support provided by the operational centers is too coarse for the needs of the US Army.

In addition, there is a need to assimilate locally collected data to improve model initialization and forecast accuracy for high-resolution nowcasts. Some previous studies using WRF-ARW for this same application (e.g., Dumais et al. 2013) have focused on using a 9-3-1-km nested grid configuration with WRF-ARW initialized using the National Center for Environmental Prediction North American Mesoscale model. Other studies (Reen et al. 2014a, 2014b) have used the Global Forecast System (GFS) model output for initialization, similar to what was used to initialize WRF-ARW for the Geographic Information System (GIS) studies to assist in the evaluation of WRF-ARW forecasts (Smith et al. 2016a, 2016b). The current study uses the GFS model for initialization.

For the model to be useful for nowcasting, a model cycle must be completed in reasonable time with the limited computational resources (multicore laptop or desktop unit in a forward-deployed location). To accomplish this it may be advantageous to perform a majority of the preprocessing and coarse model computations at a fixed-base location rather than in the field, communicate that to the field to allow the high-resolution modeling to be run, and then potentially communicate the high-resolution forecast back to the operational center. Although it is not problematic to produce the boundary conditions for the finer grids at a forward-deployed location, data assimilation of large-scale, high-volume observations such as satellite observations would be easier to accomplish at an operational center due to the bandwidth and processing power available. The constraints in such a design are then limited to the ability to efficiently communicate the data through data links to the forward-deployed location.

The forward-deployed model has the advantage that it can assimilate observations that were not yet made (or at least not yet available) when the most recent, coarser grid model was initialized (since the coarser grid model may only be run every 6 h). Another potential advantage of forward-deployed modeling is the assimilation of local observations that may not be easily or in a timely manner pushed back up

to the operational center. Since some operational centers routinely produce forecasts using increasingly finer grids, including the US Air Force's 557th Weather Wing, supplier of weather data for many US Army applications, it would then be necessary to only run the finest-scale grids in the forward-deployed environment to meet Army requirements. The goal then becomes assimilating any local initialization data and running 1 or perhaps 2 grids in a timely manner at the field location such that the forecasts are completed in a rapid update cycle mode.

One weakness of this approach is the potentially limited bandwidth for transmission of fine-resolution model data at sufficiently high temporal resolution to provide adequate boundary conditions for the 1-km (and potentially smaller) domains. Running both nests locally can have Weather Research and Forecasting Model (WRF) providing continually updated initial conditions rather than temporally interpolating between 2 initial conditions. Also, this limits the boundary conditions of the 1-km domain to account for new observations only every 6 h (the time between cycles of the coarser domains).

Although not a specific focus of this paper, the appropriate treatment of nest lateral boundary conditions will increasingly become a more important consideration for such minimal nesting Weather Running Estimate–Nowcast (WRE-N) solutions, so as to minimize known issues of limited-area mesoscale modeling like boundary sweeping (Warner et al. 1997).

1.1 Research Objectives

There are a number of goals we want to realize in this research. The ultimate goal is to deploy WRF in a forward location and produce timely and useful 0- to 3-h and perhaps 0- to 6-h forecasts (nowcasts) tailored to the individual end user. This is not a trivial task given the potentially limited communications and computational resources available in the field. Model initialization and boundary condition datasets can be large and need to be assimilated into the modeling system.

Secondly, we need to increase the resolution so that user needs are met. Specifically, we require a grid spacing of roughly 1 km to resolve approximately 5-km-scale atmospheric phenomena that are necessary to resolve the detailed flow field over complex terrain. For example, on a 300-m inner-nested grid we can resolve phenomena with a scale of approximately 1.5 km, and on a 200-m grid we can resolve to an approximately 1-km scale. To be practical, the extent of the inner grid must be limited to remain within the computational time constraints of the system being deployed in the field. The sub-kilometer-scale work is not covered in this report.

Thirdly, there is a need to provide a measure of forecast uncertainty. Since WRE-N system operators are unlikely to have a meteorological background, the need to express forecast confidence is essential. There are a number of ways to achieve this. One is to use a time-lagged ensemble (Lu et al. 2007), which involves setting up and running the model in a rapid update cycle mode and compiling statistics and variances based on sequential model output. While this is relatively easy to achieve, there is a serious limitation in that earlier forecast cycles will not have previous forecasts to compare against in deriving the statistics. The magnitude of the perturbation in a time-lagged approach depends on the age of a forecast since forecast quality usually decreases with lead time; thus, ensemble spread in a time-lagged system may not be as useful as if generated by initial conditions, stochastic, or physics-based methods. The time-lagged approach's big advantage is in ease of production and computational efficiency.

Another method to achieve the goal of developing a measure of uncertainty involves using physics-based ensembles (Stensrud et al. 2000). This method involves making several model runs with different physics packages selected and calculating model ensemble statistics from the model runs. The variety of model physical schemes incorporated into today's mesoscale models is described in texts such as Stensrud (2007). In an optimal environment this would involve developing an ensemble of 20 or more members, just as the operational centers employ. A sampling of these is described by Eckel et al. (2010). Currently, this is not practical in a field-deployed environment, but we can employ a limited-member ensemble to obtain a rudimentary measure of the uncertainty due to the model physics. If we resort to a time-lagged ensemble, it is evident that we need to select the optimal combination of physics packages for that environment. Because we do not have advance notice where the system will be deployed, we cannot fine-tune the model physics in advance for a specific environment. The goal then becomes selecting a single configuration that works well in most, if not all possible environments. Adjustments to configuration are not practical with limited lead time before deployment and, in some cases, the previously unknown location(s) where the modeling system is actually deployed.

The goal of this aspect of the research is to examine the performance of a number of physics packages to see which performs best in a complex terrain environment. We selected a southern California domain that has been widely used and documented at ARL (Dumais et al. 2009, 2013; Dyer et al. 2015, 2016; Foley et al. 2015). The inner nests of this domain include coastal, transitional, mountainous, marine, urban, agricultural, forested, and desert environments. To focus on and examine the verification of surface and near-surface properties and focus on the diurnal variation of the convective boundary layer (BL) during weak synoptic

forcing, we examined a number of planetary BL (PBL) options using the latest version of WRF-ARW (v.3.8.1, when this research began).

1.2 Common Model Configurations: Subgrid-Scale Physics

Some of the common physics scheme configurations for WRF are listed in Table 1. Included are 3 configurations used at ARL, 2 used by the 557th Weather Wing, and 1 configuration used by the National Severe Storms Lab (NSSL) for operational prediction of severe weather events over the continental United States. Some physics packages are preferred in the operational community. The NOAH (National Center for Environmental Prediction [NCEP], Oregon State, Air Force, Hydrology Laboratory) land surface model (LSM) has been almost universally accepted because of its performance and capabilities. Likewise, the combination of Dudhia and Rapid Radiative Transfer Model (RRTM) radiation schemes, Kain-Fritsch cumulus parameterization scheme, and WRF single-moment-5-class microphysics schemes are commonly chosen because of both performance and extensive testing. Note that, generally, no cumulus parameterization has been applied when using grid spacing below approximately 4 km because the cloud elements become partly resolved at that scale. More recently, however, scale-aware cumulus parameterizations have become available that can be applied to any resolution domain.

Table 1 Common physics scheme/option configurations used by ARL, Air Force, and NSSL as of this writing

Configuration	PBL scheme	SL scheme	LSM scheme	Radiation scheme	Microphysics scheme	Cumulus parameterization
ARL (Reen and Dumais 2014)	MYJ scheme (2)	Eta similarity scheme (2)	Unified NOAH LSM (2)	Dudhia/RRTM schemes (4)	WRF single-moment-5 class (4)	Kain-Fritsch Scheme (1)
ARL (Dumais et al. 2013)	YSU scheme (1)	M-O similarity (1)	NOAH LSM (2)	Dudhia/RRTM (4)	WRF single-moment-5 class (4)	None
ARL (Reen and Dumais 2018)	MYJ scheme (2)	Eta similarity scheme (2)	NOAH LSM (2)	Dudhia/RRTM (4)	Thompson microphysics (8)	Kain-Fritsch Scheme (1)
AFWA operational (DTC 2015)	YSU scheme (1)	M-O similarity (1)	NOAH LSM (2)	Dudhia/RRTM (4)	WRF single-moment-5 class (4)	Kain-Fritsch Scheme (1)
AFWA alternative test configuration (DTC 2015)	ACM2 PBL scheme (7)	Pleim-Xiu scheme (7)	Pleim-Xiu LSM (7)	Dudhia/RRTM (4)	WRF single-moment-5 class (4)	Kain-Fritsch Scheme (1)

Table 1 Common physics scheme/option configurations used by ARL, Air Force, and NSSL as of this writing (continued)

Configuration	PBL scheme	SL scheme	LSM scheme	Radiation scheme	Microphysics scheme	Cumulus parameterization
NSSL operational WRF	MYJ scheme (2)	MYJ scheme (2)	NOAH LSM (2)	Dudhia/RRTM (4)	WRF single-moment-6 class (6)	None (4-km grid)
High resolution rapid refresh (HRRR)	MYNN 2.5 (5)	MYNN (5)	Rapid Update Cycle (3)	RRTMG (4)	Thompson aerosol aware (28)	None

Notes: M-O = Monin-Obukhov; MYJ = Mellor-Yamada-Janjic; YSU = Yonsei University; ACM2 = Asymmetric Convection Model 2; MYNN = Mellor-Yamada-Nakanishi-Niino; the G in RRTMG stands for Global Climate Model; AFWA = Air Force Weather Agency (now 557th Weather Wing), which is discontinuing use of the WRF model in favor of the Global Air Land Weather Exploitation Model (GALWEM) model, which is based on the UK Meteorological Office model; HRRR may be found at <https://rapidrefresh.noaa.gov/hrr/HRRR/static/HRRRv3-ESRL/wrf.nl>; numbers in parentheses () are option numbers in the WRF model for the type of scheme, and represent switches corresponding to the particular physics scheme in the mode.

Compared with some of the other physical parameterizations included in WRF, there seems to be greater diversity in the user community in which PBL/surface layer (SL) physics combination is chosen. While the MYJ and YSU schemes are both considered “favorites”, the operational models represented in the table suggest that this is not universal. In addition, the BL schemes may have multiple options for the SL choices, but in the present study we have chosen to couple the most-common intended SL scheme to a particular BL parameterization choice. Since the SL and BL schemes have been designed to work with each other, they are considered optimal choices, and we adhere to this criterion here.

Figure 1 shows the spatial regimes for which the physical parameterizations used in mesoscale models may not be strictly valid. A parameterized BL assumes all large turbules are subgrid scale and can be treated in an ensemble sense. This is similar to deep cumulus convective parameterizations (also known as convective parameterizations) originally employed in mesoscale models for grid spacing greater than 10 km, where they are necessary to properly represent the release of latent heat since the convective eddies are assumed to be entirely subgrid scale. This assumption breaks down when using smaller grids, but they have been found to be helpful in triggering convection in 5- to 10-km grid applications (Skamarock 2008). Generally, these cumulus parameterizations are not used when the model is capable of resolving the convective eddies, typically at 4- to 5-km grid spacing. As an aside regarding deep cumulus convection, true deep-convection resolving (as opposed to deep-convection allowing) may not be achieved until grid spacing of near 100 m is approached.

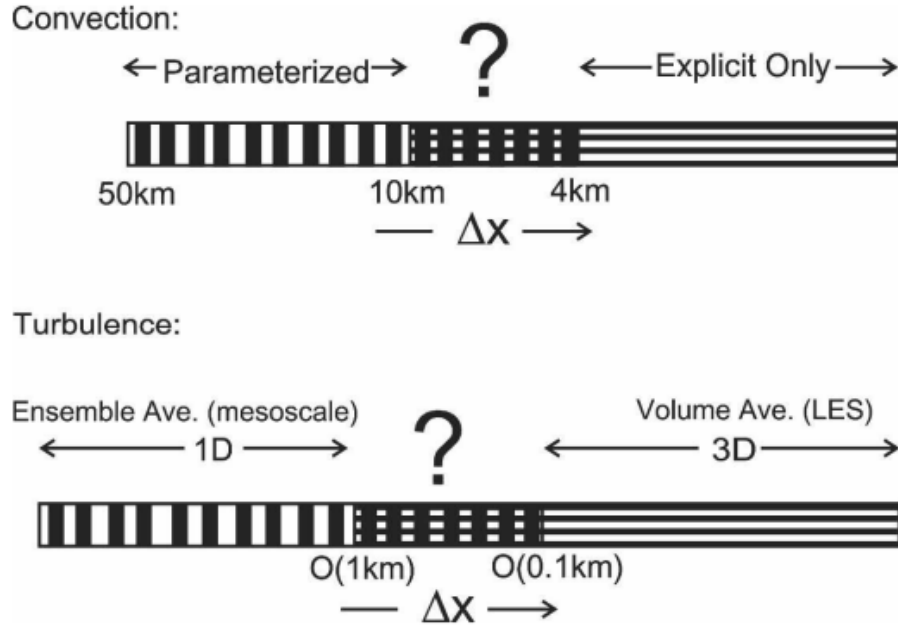


Fig. 1 Horizontal grid-scale dependence of convection and turbulence parameterizations (Deng and Stauffer 2006). Cross-hatched areas indicate where the underlying assumptions of the physical parameterizations are not strictly valid, and there is currently no robust, satisfactory solution in these no-man’s land scale ranges: convection (4–10 km Δx) and turbulence ($\Delta x \sim$ depth of mixed layer).

The larger convective BL turbulent structures and shallow convection processes, in general, start getting essentially fully resolved at scales of approximately 200 m to 1 km (Bryan et al. 2003; Wyngaard 2004; Deng and Stauffer 2006). Below approximately 100 m, large-eddy-simulation (LES) models may be used in an attempt to resolve features smaller than the depth of the PBL. However, LES models are computationally intensive and likely not practical for use in a nowcasting system with a tight time schedule within which they need to run and produce the necessary output for decision making. If we wish to resolve features of a size scale approximately the depth of the daytime convective BL, a grid spacing of approximately 200–350 m is required. A grid spacing of 1 km allows us to resolve features approximately 5 km in size, although what constitutes “resolved” may be tricky. Skamarock et al. (2008) suggests $7 \Delta x$, for instance. For the purposes of this study, the innermost, finest grid has a 1-km grid spacing. This allows us to represent a significant portion of the scales we wish to identify and represent in our simulations, including moderately fine-scale topographic-induced flows. We are confining this study to the development of the daytime convective BL where there is no moist convection, and therefore we are not examining the convective parameterizations since they parameterize moist convection. Those physics are merely turned off for the innermost (3- and 1-km) grids.

Our primary goal in this research is to determine which of the BL/SL parameterization schemes works best for generalized forward deployments where the deployed location is not known in advance. Because the modeling system is preconfigured, one scheme will be set up for use in the deployed location, regardless of location. Our modeling domain was also chosen to represent a number of geographic regions focusing on complex terrain. This goal is challenging from the standpoint of not only capturing a complex range of physical processes, land use, and elevations, but also accounting for the steep gradients in elevation, which can potentially make achieving modeling stability criteria more difficult.

1.3 Structure of Report

This report examines the performance of PBL and SL physics in mesoscale models, specifically the WRF-ARW modeling system and its use in the WRE-N configuration developed by ARL as a means of evaluating the merits of the selected BL/SL parameterization schemes. Section 2 describes the PBL and its structure, focusing on situations with little synoptic forcing. In this way, we largely consider BL physics and processes, in the absence of strong synoptic scale forcing, and emphasize the diurnal variation of the PBL. Section 3 describes the various physics options presented in the current (v.3.8.1) version of WRF-ARW. Section 4 describes the model configuration for this set of experiments. The results of this series of experiments are described in Section 5. We examine the traditional measures of model performance for our tested physical schemes, and examine the evolution of the daytime and nocturnal BL, particularly near the surface. Section 6 summarizes the conclusions. Finally, Section 7 explores some additional considerations that are needed to make the WRE-N modeling system viable given current computational constraints in light of the limitations of the current hardware and software.

2. Planetary Boundary Layer

The atmospheric BL is dynamic. It is also not well understood. There is a dearth of observational data, and what we know comes from a few well-known field experiments over prescribed terrain types. A tabulation of the more-well-known experiments that have formed the foundation of today's turbulence and BL theory is given in Stull (1988). Many of these field studies were conducted over 40 years ago, and the results still form the basis of many physical parameterization schemes. For the most part, these earlier studies were conducted over homogeneous terrain and have ignored complex terrain with the intention of being able to replicate the basic processes that control the diurnal evolution of the BL. The availability of remote sensors such as wind profilers and Light Detection and Ranging (Lidar)

have expanded our understanding of BL processes and allow us to make volume measurements and profiles rather than only point measurements, filling in the holes between measurement points (Wilczak et al. 1996). The Cooperative Atmospheric Surface Exchange Study (CASES-99) focused on specific issues related to the stable, nocturnal BL, including the transition periods (Poulos et al. 2002). One of these goals was to collect data during the transition from a convective to stable BL regime and vice versa and compare them with existing models of this transition. Most recently, the MATERHORN field experiment (Fernando et al. 2015) focused on obtaining measurements of the PBL in a complex terrain environment, but those data have not been fully exploited at the time of this writing. Many of the earlier field experiments attempted to document the structure and evolution of the BL but were extremely limited in scope partly because of limitations in measurement technologies.

2.1 Boundary Layer Development and Structure

The diurnal variation of the homogeneous BL, PBL, or atmospheric BL (ABL), all used interchangeably over land is illustrated in Fig. 2 (Stull 1988). The most-notable features are the dynamic nature and its complexity. This model of PBL structure largely ignores rapidly changing synoptic-scale effects on the structure and dynamics of the PBL.

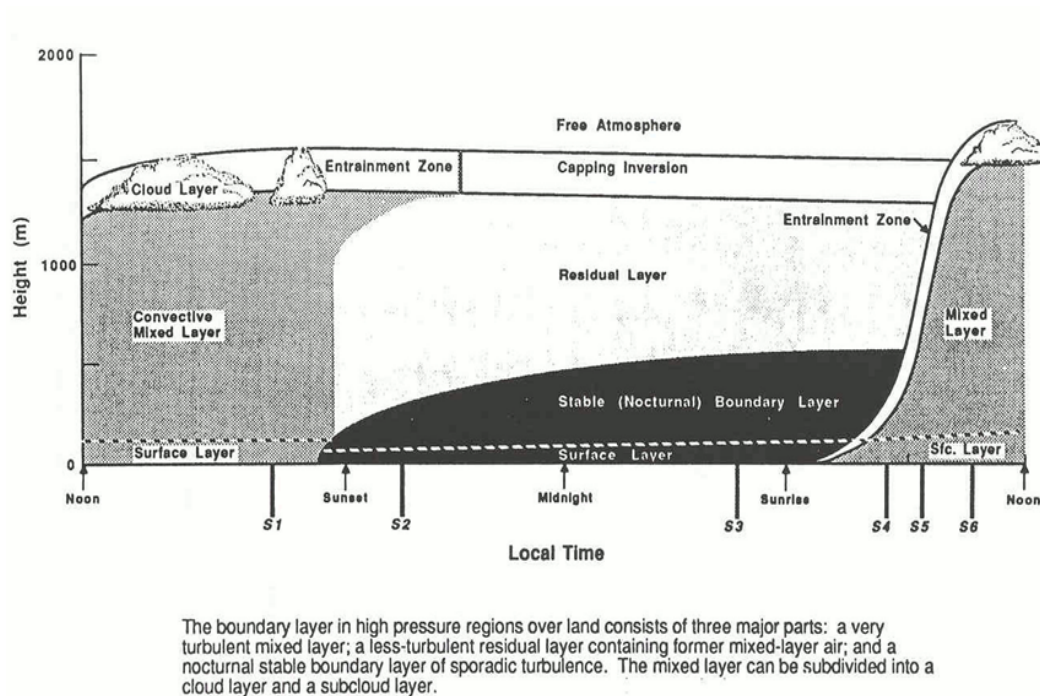


Fig. 2 Diurnal evolution and cycle in PBL structure over land and under quiescent (fair weather) conditions (Stull 1988)

Aside from this, it is easily seen that the structure of the daytime PBL is very different than the nighttime, largely stable PBL. Just after sunrise, a deep mixed layer (ML) begins to evolve, which is largely driven by the surface heating created by the absorption of solar radiation at the surface and conversion into heat. Thermals develop that mix the BL, or “convective” BL (CBL) (hence, the term “mixed layer”) until the depth reaches a maximum in the midday and afternoon. If adequate moisture is present, the ML is capped by a cloud layer. Thermals overshoot into the comparatively overlying stable free atmosphere/air (FA), as air from above the capping inversion is mixed into, or entrained into, the ML. Under the influence of a typical high-pressure system, there is normally a subsidence inversion capping the ML. Despite the stability, stronger thermals are able to penetrate into the FA and mix some of it downward into the ML, causing a deepening of the ML. The ML actually begins entraining air from above slightly after sunrise, and this process continues into late afternoon.

At nighttime, the BL is quite different. As the surface cools radiatively, the air above it cools, resulting in the development of a stable BL (SBL). Above the SBL lies the residual layer, which is the remnant of the daytime CBL and where only episodic turbulence occurs. It is clear that correctly representing physical processes at the surface is also very important; thus, the appropriate choice of a surface formulation in numerical simulations is necessary. Correct parameterization of radiative processes is also very important. In addition to the development of an SBL, an SL develops in direct contact with the surface. The SL is usually very thin, hence the need for very high vertical resolution in numerical models near the surface. The SL is defined as that part of the BL where turbulent fluxes and stress vary by less than 10%, hence this is typically defined as the lower 10% of the BL, whether it is the CBL or SBL, and independent of time of day.

What should be evident is that the BL structure and evolution is highly complex. The nocturnal SBL, for example, is extremely difficult to parameterize due to the complexity of interacting (and intermittent) subgrid-scale wave and turbulent processes and extreme near-surface gradients of atmospheric parameters (Seaman et al. 2012). Critical modeling aspects that need to be accounted for include the influence of synoptic scale weather patterns, inhomogeneity of the underlying surface, complex topography, and instances where the surface happens to be water. Aside from these considerations, the BL is also complicated near transition regions, for example, from a water surface to land. Simulations of such complex environments (Smith et al. 2016a, 2016b) have attempted to bin performance statistics on the basis of distance from the shoreline, elevation, topography, and latitude, and these subdomains can produce quite different statistics.

2.2 Features within the Convective Boundary Layer

The largest PBL structures have dimensions approximately equal to the depth of the PBL (Stull 1988). Examples of these larger structures include both plumes and cloud streets. Small-scale flows can arise in complex topography, and the local winds that develop can have a significant impact on battlefield operations, whether it be for artillery operations or for flight plans involving drones.

Plumes arise when large convective structures develop from inhomogeneous surface heating (Fig. 3). They are commonly present in nearly all daytime convective BLs (Kaimal and Businger 1970; Wilczak and Tillman 1980). The vertical dimension of plumes is limited by the depth of the daytime PBL, typically 1–2 km, and the horizontal dimension is roughly the depth of the SL, approximately 100 m (Stull 1988). Properly resolving these features, at least in the vertical, requires a vertical grid spacing of a few hundred meters or less. While LESs are available and integrated within the WRF model infrastructure, the ability to resolve these features is computationally intensive. Resolving these features in the horizontal requires at least 5 grid points, which implies model grid spacing of a few hundred meters to approximately 1 km.

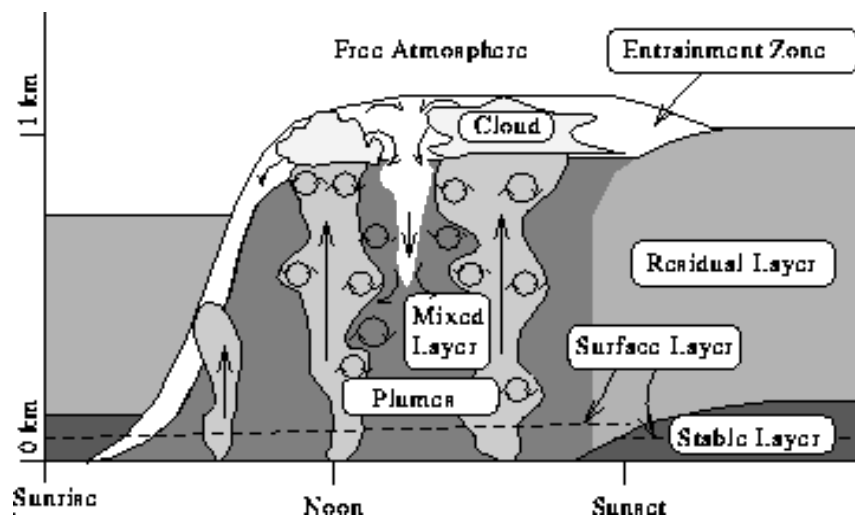


Fig. 3 BL plume structures. The vertical dimension is determined by the depth of the BL, and is typically equal to or larger than the horizontal dimension of the plumes. Smaller-scale turbulent motions exist within the cloud structure (Reproduced with permission from http://lidar.ssec.wisc.edu/papers/akp_thes/node6.htm).

Under certain conditions of shear and stability, horizontal roll circulations may be present (Etling and Brown 1993). Roll circulations are visible in satellite imagery where sufficient moisture is present, forming cloud streets. A diagram of these rolls is shown in Fig. 4. Cloud streets form along and at a small angle to the geostrophic

wind. The circulations that develop, including alternating updraft and downdraft circulations, are superimposed on the large-scale flow. The vertical depth of these circulations is on the same order as plumes, limited by the depth of the PBL. Diurnal variation in the depth of the PBL occurs and modulates the formation of these circulations. The horizontal dimensions of these rolls frequently exceed the depth of the PBL, the smallest horizontal dimensions being approximately the depth of the PBL. Roll circulations are readily observed in satellite imagery, particularly when cold air advects over open warm water, a good example being those that occur during winter off the east coast of the United States over the warm Gulf Stream. While this example is the most easily observed, cloud streets also form over land, and the circulations may be present in drier environments that are incapable of supporting cloud formation.

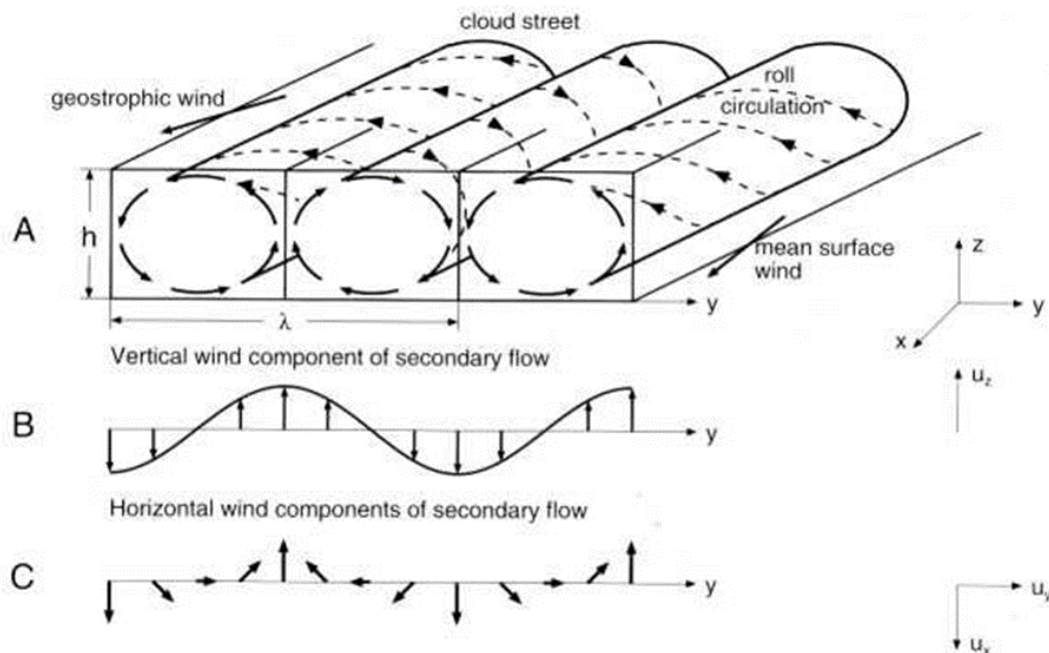


Fig. 4 Structure of horizontal BL rolls, showing the circulations that develop and associated sinusoidal behavior of the vertical motion field. The wavelength of these rolls is approximately 2–5 km. Cloud streets form in sufficiently moist BLs in the upward circulation between rolls. (Alpers and Brummer 1994).

In addition to PBL features, topographically induced circulations may develop within the PBL that have significant impact on battlefield operations. Mountainous terrain must be represented within these models by the use of high-resolution topography and land-use information commensurate with the grid scale chosen for each nested domain.

2.3 Sources of Model Error

Forecast error related to numerical weather prediction (NWP) is a continuing topic of discussion in numerical modeling textbooks (e.g., Warner 2011) and in the published literature. Stauffer (2013) explores the uncertainty in environmental NWP modeling. At a high level, errors can be grouped into 2 general categories: model error and measurement error. Since both observations and models contain information representing different scales, this contributes to a degree of uncertainty. We need to consider the sampling, what both the model and the instruments are representing, and over what time and spatial scales. Rarely do these match. If we look at modeling error, there are 3 sources of error: initial and lateral boundary conditions, numerics, and physics. Models also need to “spin up” so that the simulated atmosphere is allowed a chance to achieve balance (i.e., adjusting to equilibrium) and thus create the circulations that respond to the initial conditions supplied, and the topography. This problem of ingesting local observations and allowing spin-up is mitigated in numerical weather prediction by applying a common data assimilation approach through a portion of the forecast—“observation nudging” (Liu et al 2005; Deng et al. 2009), or relaxing the model gradually toward the observations. In our research we incorporated the observational nudging scheme built into WRF called Four-Dimensional Data Assimilation (FDDA) (Reen 2016) the initial 6 h of the forecast. It is not applied at a single time; rather, it is a continuous form of data assimilation since it is applied at every time step over a specified period. Nudging is computationally efficient, continuous, and adaptable to fine grids such as those used here, although nudging cannot as readily handle indirect observations of weather parameter as variational data assimilation approaches.

Lateral boundary errors can also arise due to advection from outside the grid where initial conditions are not prescribed due to lack of data. Hence, the nesting used in high-resolution models must account for the propagation of meteorological (and other physical) properties from outside the nest under consideration. Meteorological phenomena typically propagate with speeds of 5–50 ms^{-1} , but external gravity waves can travel with speeds as high as 300 ms^{-1} , which controls how quickly errors introduced at the lateral boundary conditions can sweep through the domain. In addition, it is advisable to avoid placing lateral boundaries in regions of strong topography (e.g., mountains) because of the strong vertical stratification of the atmosphere compared with the horizontal (Warner et al. 1997). Because rapidly varying topography results in large gradients and accelerations that are not treated realistically by the lateral boundary conditions, the inertia–gravity waves produced by misrepresenting the effects of this forcing can propagate rapidly to the

domain interior. This is also one factor limiting the nesting of finer vertical grids in mesoscale models.

Aside from the representativeness in the observations land use/soil properties and their initial conditions, and potential quality control issues, recent studies have identified a potential geolocation issue with the WRF modeling system (Monaghan et al. 2013). This issue could be more problematic with the use of high-resolution configurations, as the modeling system is now regularly used at 1- to 10-km grid spacing. WRF uses a spherical coordinate system, as opposed to the more accurate assumption of a spheroidal shape. This can introduce significant error in the mapping of geographic data, as great as 10–20 km, which offsets the data by several grid points when using 3- or 1-km grid nests like those used here. This problem was reported by Monaghan et al. (2013), and the issue came up in earlier analyses performed at ARL by Foley et al. (2015) and Smith et al. (2016a, 2016b). Monaghan et al. (2013) state that in the mid-latitudes the location of a point can vary between spheroidal and spherical earth models by more than 20 km.

In studies involving GIS we use both a projection and a datum. A projection is a mathematical formula for flattening a 3-D earth onto the flat surface of paper or a computer screen. There are 3 qualities that are important for a projection: shape, area, and distance. A projection can preserve only 1 or 2 of these but never all 3. The purpose of a datum is to compensate for the fact that the earth is not a perfect sphere but is rather oblate with a somewhat “lumpy” (uneven) surface. In the United States, North American Datum of 1983 (NAD83) has been a very common datum and the datum used in GIS-related studies at White Sands Missile Range, New Mexico. The World Geodetic System 1984 (WGS84) was constructed with satellite measurements and is very similar to NAD83. Error could potentially be introduced when PointStat matches up the WRF surface that has a conic conformal projection and no datum-to-surface observations in NAD83 datum (even though no projection/datum is stated with the surface observations). When the WRF surface was imported directly into GIS, the spherical, no-datum orientation was converted to NAD83 datum with assigned projection and the data geolocated properly (Foley 2015).

In terms of the actual model, as long as all the input data sources are on WGS84 with no attempt to convert to a sphere, although WRF may be off by 20 km in latitude since the input data sources are shifted by that same amount, everything lines up with itself. The only exceptions to this would include 1) when the input coarse grid model data (e.g., GFS) may also be assuming a spherical earth and thus not consistent with the topography (the coarseness of the GFS data used should minimize the impact of this shift) and 2) when latitude-dependent processes have errors; for example, the Coriolis effect and sun angle (the difference in the strength

of Coriolis 20 km apart minimal compared with other forcings, as would the differences caused by the variation in sun angle). Thus, it is not believed that this represents a significant source of error in our analyses.

We also need to consider the validity of various modeling assumptions. For example, many of the current PBL and SL schemes have been developed when model resolution was typically much coarser than present day modeling systems. Others are much newer. Some of the assumptions that went into developing these formulations may no longer be valid with smaller (sub-kilometer) grid spacing. The Shin-Hong (SH) scheme was developed to address the potential issue of scale awareness when sub-kilometer-scale grids are used. While many of the assumptions used in these schemes may still be fine for the stable nocturnal BL, the assumptions may not hold for the deeper, daytime convective PBL where $\Delta \sim \ell$, where Δ represents the grid spacing and ℓ is the length scale of the BL, which is largely controlled by the BL depth.

This is not an all-inclusive discussion of the sources of model error and model verification. However, it does shed light on the challenges we face in assessing model performance and the sources of error most applicable to the current study. While the fundamental sources of uncertainty in mesoscale modeling have been discussed, neither has every source of model error been considered and incorporated into the analysis, nor is it possible to completely address all of the issues involving the introduction of model error.

3. WRF Boundary-Layer/Surface-Layer Physics Packages

A variety of PBL schemes have been included with WRF since its inception (Pagowski 2004). These schemes have evolved in complexity and number since the initial version of the model. WRF v3.8.1 contains 13 distinct, selectable PBL schemes, and for the most part each is paired with at least 1 SL scheme. Some PBL options allow the choice of more than one coupled SL scheme. In addition to the PBL/SL schemes, WRF allows the use of an LES model that may be used for finer-grid spacing simulations. Because of the computational expense associated with running WRF at the resolutions necessary for LES, we have limited our finest-grid spacing to 1 km and thus have confined our study to the traditional PBL/SL schemes.

3.1 WRF v.3.8.1 Boundary-Layer/Surface-Layer Schemes

WRF contains a number of user selectable BL and SL schemes. Table 2 lists the available PBL/SL schemes and combinations within the WRF v3.8.1 modeling system framework. Noninclusive of the aforementioned LES option, there are 13

options for the PBL. Each PBL option is coupled to one or more SL scheme(s). The table also gives the year of the inclusion into WRF, and lists the prognostic and diagnostic variables associated with each scheme (where applicable). The final column indicates whether it is a local, nonlocal, or hybrid scheme. The latter is important when it comes to understanding how the scheme has been designed to represent the gradients associated with the internal calculations and gives insight as to what eddy size the scheme is best suited to represent. Local schemes calculate the gradient by using adjacent vertical grid points and thus may represent the smaller-scale eddies and their effects better. Nonlocal schemes typically are expected to perform better in capturing physical processes associated with the larger BL eddies. While hybrid schemes may be expected to perform well with either eddy size, this is not necessarily always the case. Generically then, PBL schemes may be categorized into 3 categories based on this criteria.

Table 2 Planetary BL and SL options available in WRF v3.8.1 including pairing information, prognostic and diagnostic variables available in each scheme, and whether it is a local, nonlocal, or hybrid scheme

PBL scheme	WRF option	Year intro.	SL pairing	Prog. var.	Diag. var.	Local/nonlocal/hybrid
Yonsei University Scheme (YSU) (Hong et al. 2006)	1	2004	1,91	...	exch_h	N
Mellor-Yamada-Janjic Scheme (Janjic 1994)	2	2000	2	TKE_PBL	el_myj, exch_h	L
NCEP Global Forecast System Scheme (Hong and Pan 1996)	3	2005	3	TKE_PBL	el_pbl, exch_h, exch_m	...
Quasi-normal Scale Elimination (QNSE) Scheme (Sukoriansky et al. 2005)	4	2009	4	QKE	Tsq, Qsq, Cov, exch_h, exch_m	L
Mellor-Yamada-Nakanishi-Niino (MYNN) Level 2.5 and Level 3 Schemes (Nakanishi and Niino 2006, 2009)	5, 6	2009	1,2,5,91	QKE, Tsq, Qsq, Cov	exch_h, exch_m	L
Asymmetric Convection Model 2 Scheme (ACM2) (Pleim 2007a)	7	2008	1,7,91	H
Bougeault-Lacarrère Scheme (BouLac) (Bougeault and Lacarrère 1989)	8	2008	1,2,91	TKE_PBL	el_pbl, exch_h, exch_m, wu_tur, wv_tur, wt_tur, wq_tur	L
University of Washington (turbulent kinetic energy (TKE) Boundary Layer Scheme (Bretherton and Park 2009)	9	2011	1,2,91	TKE_PBL	exch_h, exch_m	L

Table 2 Planetary BL and SL options available in WRF v3.8.1 including pairing information, prognostic and diagnostic variables available in each scheme, and whether it is a local, nonlocal, or hybrid scheme (continued)

PBL scheme	WRF option	Year intro.	SL pairing	Prog. var.	Diag. var.	Local/nonlocal/hybrid
Total Energy Mass Flux (TEMF) Surface Layer Scheme (Angevine et al. 2010)	10	2011	10	TE_TEMF	*_temf	H
Shin-Hong Scale-aware Scheme (Shin and Hong 2015)	11	2013	1,91	...	exch_h, tke_diag	H
Grenier-Bretherton-McCaa Scheme (Grenier and Bretherton 2001)	12	2015	1,91	TKE_PBL	el_pbl, exch_tke	L
Medium Range Forecast Scheme (Hong and Pan 1996)	99	2000	1,91	N

Cohen et al. (2015) summarized the characteristics of a number of schemes from WRF v3.3.1. Their detailed analysis involved a cold-season severe convective situation over the southeastern US. Since v3.3.1, a number of schemes and enhancements have been added to the model. Since WRF is not a static modeling system, the available packages are updated, and some removed and replaced, on a regular basis. We describe the schemes available in WRF v.3.8.1 here. While Cohen et al. (2015) is not the only comparative study performed on the WRF PBL schemes, it is one of the newest and most comprehensive. As is the case with the other reviewed studies, Cohen et al.’s study applies to one specific environment and set of synoptic conditions.

3.1.1 Yonsei University Scheme (YSU)

The YSU (Hong et al. 2006) is a nonlocal first-order closure scheme that is similar to the Medium Range Forecast model (MRF) scheme, except it explicitly represents entrainment at the top of the PBL. Its origins are from the MRF scheme. The original MRF scheme evolved into the YSU and GFS schemes. Later, the SH scheme evolved from the YSU scheme.

The YSU scheme was developed as an improved replacement in WRF for the MRF scheme. It is a nonlocal-K (eddy diffusivity) scheme with an explicit entrainment layer and a parabolic K profile in unstable mixed layers. It includes a topographic correction for surface winds to represent extra drag from subgrid topography and enhanced flow at hill tops (Jimenez and Dudhia 2013). The topographic correction option works with the YSU PBL only, and it was introduced in WRF v3.4. Further improvements to the YSU scheme include a simpler terrain-variance-related

correction and an option for top-down mixing driven by radiative cooling. While the scheme more accurately simulates deeper vertical mixing in buoyancy-driven PBLs with shallower mixing in strong-wind regimes compared with MRF (Hong et al. 2006), it still has been found to over-deepen the PBL for springtime deep convective environments. This results in too much dry air near the surface and underestimation of mean layer convective available potential energy related to deep convection (Coniglio et al. 2013).

3.1.2 Mellor-Yamada-Janjic Scheme (MYJ)

The MYJ Scheme (Janjic 1994) is a local 1.5-order closure scheme that includes a prognostic equation for turbulent kinetic energy (TKE). It was the operational scheme used in the Eta model and includes a 1-D prognostic TKE scheme with local vertical mixing. While this scheme improves upon the original Mellor-Yamada 1.5-order local scheme (Mellor and Yamada 1974, 1982), without particularly large computational expense, it under-mixes PBL for locations upstream of spring convection (e.g., Coniglio et al. 2013).

3.1.3 National Center for Environmental Prediction (NCEP) Global Forecast System Scheme

The NCEP Global Forecast System Scheme (Hong and Pan 1996) is available for the WRF Nonhydrostatic Mesoscale Model version only. It is the scheme currently used in NCEP's operational numerical weather prediction model GFS. Since this scheme is not available for use with WRF-ARW, it has not been evaluated in this study.

3.1.4 Quasi-normal Scale Elimination (QNSE) Scheme

Similarly to the MYJ scheme, the QNSE scheme (Sukoriansky et al. 2005, 2008) is a local 1.5-order local-closure scheme designed to account for wave phenomena within stable BLs. It includes a TKE-prediction option that uses a novel formulation for stably stratified regions and has been an option in WRF since v3.1. In the daytime CBL, the scheme uses an eddy diffusivity mass-flux method with shallow convection added in v3.4. QNSE was found to provide a realistic depiction of potential temperature profiles, PBL depth, and kinematic profiles based on comparison with observational data and corresponding large eddy simulations (Kosovic and Curry 2000) under stable conditions. However, as with the MYJ scheme, in the case of less-stable PBLs, QNSE results in too cool, moist, and shallow PBLs in simulations of springtime convective environments (Cohen et al. 2015).

3.1.5 Mellor-Yamada Nakanishi Niino (MYNN) Levels 2.5 and 3 Schemes

The MYNN Levels 2.5 and 3 Schemes (Nakanishi and Niino 2006) include 2 options: a 1.5-order (MYNN2) and second-order (MYNN3) closure scheme. Like the QNSE scheme, MYNN is a local scheme based on results from comparison with large eddy simulations. In contrast, the original Mellor-Yamada PBL scheme (Mellor and Yamada 1974, 1982) on which this scheme is based expresses stability and mixing length and is based on comparison with observations. The expressions for mixing length with MYNN are applicable to a greater variety of static stability regimes than the original formulation (Nakanishi and Niino 2009). The MYNN Level 2.5 PBL predicts subgrid TKE terms and first appeared in WRF v3.1, with significant updates in v3.8. Subgrid-scale clouds from MYNN are coupled with radiation, and cloud water and ice mixing is included. These options are new in v3.8. MYNN2 is less computationally expensive than MYNN3, but in addition to the improvements with MYNN 2.5, MYNN3 better simulates deeper mixed layers than the Mellor-Yamada PBL scheme (Mellor and Yamada 1974, 1982), and MYNN3 reasonably represents statically stable BLs in the case of radiative fog development (Nakanishi and Niino 2006). MYNN2 improves the PBL over nonlocal PBL schemes for springtime PBLs that support deep convection (Coniglio et al. 2013). However, as with the MYJ scheme, the local formulations of both MYNN2 and MYNN3 still may not fully account for deeper vertical mixing associated with larger eddies and associated countergradient flux.

3.1.6 Asymmetric Convection Model 2 Scheme (ACM2)

The ACM2 Scheme (Pleim 2007b) is a first-order closure hybrid scheme where upward fluxes within the PBL are represented as interactions between the SL and every layer above (including local eddy diffusion), while downward fluxes extend from each layer to each immediately underneath. Convective plumes in the PBL arise from the diurnally heated SL, whereas downward fluxes are more gradual. In the ACM2 PBL we have nonlocal upward mixing and local downward mixing. This scheme was added in WRF v3.0 (Pleim 2007a) and asserts that the PBL potential temperature and wind velocity profiles are depicted with greater accuracy when both local and nonlocal viewpoints are considered regarding vertical mixing. Pleim (2007b) further validates the use of the ACM2 scheme because PBL depths produced by the scheme are similar to those derived from afternoon wind profiler data. However, Coniglio et al. (2013) find errors for evening soundings supporting deep convection with ACM2 over-predicting the depth of the PBL.

3.1.7 Bougeault–Lacarrère Scheme (BouLac)

The BouLac Scheme (Bougeault and Lacarrère 1989) is a 1.5-order local closure scheme including a prognostic equation for TKE. Its design is most relevant for terrain-enhanced turbulence (e.g., wave phenomena) and its impact on the PBL. The BouLac was first incorporated in WRF v3.1 and specifically designed for use with the Building Effect Parameterization urban model (Salamanca et al. 2011). It was found to improve representation of the PBL in regions of higher static stability compared with nonlocal schemes in similar environments (Shin and Hong 2011).

3.1.8 University of Washington (UW) TKE Boundary-Layer Scheme

The UW Boundary Layer Scheme (Bretherton and Park 2009) is a 1.5-order local closure scheme that attempts to improve upon the Grenier-Bretherton-McCaa model including accounting for relatively longer time steps (relevant for climate models), diagnosing rather than forecasting TKE and focusing computations over a number of layers determined by the vertically varying stability of the thermodynamic profile. The UW scheme is from the Community Earth System Model (Hurrell et al. 2013) and was introduced in WRF v.3.3. While UW was developed to provide a better depiction of the nighttime stable BL than existing PBL schemes at the time, substratocumulus layers develop that are too moist relative to results from large eddy simulations (Holtslag and Boville 1993). Nonlocal PBLs are less moist and well mixed using the UW scheme than LES models, which are too moist. This scheme has similar disadvantages as with the MYNN and BouLac schemes, meaning it does not adequately address the effects of larger eddies and the counter-gradient corrections associated with these larger eddies.

3.1.9 Total Energy Mass Flux (TEMF) Surface Layer Scheme

The TEMF scheme (Angevine et al. 2010) is a 1.5-order closure hybrid scheme including a nonlocal component represented by updrafts triggered by upward heat fluxes originating at the surface, creating mass fluxes throughout the PBL by applying a counter-gradient correction (Angevine 2006). The scheme is summarized by Cohen et al. (2015). The local component is triggered under statically stable conditions using a total turbulent energy concept that eliminates buoyant destruction of TKE during high static stability (Mauritsen et al. 2007). The scheme includes a subgrid total energy prognostic variable plus mass-flux-type shallow convection. TEMF was incorporated in WRF v.3.3. The scheme compared favorably with LES results for observations from the 2006 Texas Air Quality and Gulf of Mexico Atmospheric Composition and Climate Study (Texas Air Quality Study II/Gulf of Mexico Atmospheric Composition and Climate Study) around

Houston, Texas (Angevine et al. 2010). It gives improved PBL profiles with shallow cumulus clouds compared with the LES models (Angevine et al. 2010). However, the scheme produces excess drying beneath BL stratocumulus clouds and is moister within the lower cloud layer than results from the LES models, indicating excessive moisture flux across the lower cloud boundary in the TEMF scheme (Angevine et al. 2010).

3.1.10 Shin-Hong Scale-Aware Scheme

The SH Scale-Aware Scheme (Shin and Hong 2015) is relatively new in WRF, first appearing in WRF v.3.7. This hybrid scheme incorporates a scale dependency for vertical transport in the convective PBL. Vertical mixing in the stable PBL and free atmosphere follows YSU. The main improvements in the SH PBL scheme include prescribing the nonlocal heat transport profile fitted to the output from an LES model and the inclusion of an explicit scale-dependency function for vertical transport in CBL. Thus, the SH PBL parameterization is designed to work at the gray zone (100 m to 1 km) resolution for the CBL. Subgrid-scale transport is described by Shin and Hong (2013). Nonlocal transport (large eddies) and local transport (smaller eddies) are calculated separately. Subgrid-scale transport is multiplied by a grid-scale-dependency function, and the local transport is then calculated through an eddy diffusivity formulation containing both subgrid scale and local terms. The grid size dependency is in the form of an empirical function fitted to reference data from a free convection case. This scheme also has diagnosed TKE and mixing length as output.

3.1.11 Grenier-Bretherton-McCaa Scheme

The Grenier-Bretherton-McCaa Scheme (Grenier and Bretherton 2001) is a local TKE scheme tested in cloud-topped PBL cases. This scheme combines a level-1.5 turbulent closure model with an entrainment closure at the top of the PBL. The model performs very well in cases of dry convection and also performs well in stratocumulus-topped BLs with models configured with 15 hPa or better vertical resolution. The authors of this scheme recommend that for comprehensive simulation of BL convection this scheme be coupled with a shallow cumulus parameterization scheme. This is a TKE scheme and first appears in WRF v.3.5.

3.1.12 Medium Range Forecast Scheme

The MRF Scheme is a first-order closure nonlocal scheme that incorporates a countergradient correction term into downgradient diffusion expressed solely by local mixing. The original formulation of the MRF scheme has implicit treatment of the entrainment layer as part of nonlocal-K mixed layer (Troen and Mahrt 1986),

and the original scheme is offered as separate option (99) in WRF v3.8.1, along with the YSU PBL option (1), which is an improved MRF scheme rewritten specifically for WRF. Compared with local PBL schemes, the MRF scheme was designed to more accurately simulate the deeper mixing within an unstable PBL where larger eddies entrain higher potential temperature air from above the PBL into the PBL (Wyngaard and Brost 1984; Stull 1993). The MRF scheme was found to depict too deep of a PBL, especially in strong wind regimes at night (Mass et al. 2002), an issue particularly troublesome in wind energy simulations and the excessive mixing results in over-erosion of convective initiation (Bright and Mullen 2002).

3.2 Other PBL Scheme Intercomparison Studies

Prior to WRF, PBL scheme investigations took place using earlier models, such as the Penn State National Center for Atmospheric Research (NCAR) Model (MM4 and MM5), as these schemes were being developed and incorporated into mesoscale models. Holt and Raman (1988) investigated first-order and TKE closure schemes and compared model statistics with observations. They found that the mean structure of the PBL is fairly insensitive to the type of closure, and that TKE closure is preferable to first-order closure in predicting the overall turbulent structure of the PBL. As LES models became formulated and available, Ayotte et al. (1996) noted weaknesses in representing entrainment and employed tuning techniques using LES data to improve simulations. In that study the PBL scheme changes were applied by actually going into the code and modifying the physical processes.

The YSU and MYJ BL schemes were evaluated for WRF with an innermost grid of 1.33-km resolution for simulations of Hurricane Isabel (Nolan et al. 2003a, 2003b). These schemes were chosen for comparison since both are widely used and are based on entirely different methods for simulating the PBL (Nolan et al. 2003a). The original YSU and MYJ schemes were modified to have ocean roughness lengths in closer agreement with observations. Increasing the horizontal resolution around the eyewall also leads to improved BLs as well as an improvement of the vertical structure of the inner-core wind field (Nolan et al. 2003b).

Hu et al. (2010) tested 3 WRF v.3.0.1 PBL schemes, namely MYJ, YSU, and ACM2, comparing observational data over eastern Texas for air quality purposes. They performed 92 runs and examined surface data and vertical profiles where available. They found that the differences between the schemes were primarily due to vertical mixing strength and entrainment from air above the PBL. They performed a sensitivity experiment to confirm this hypothesis. The ACM2 and YSU

nonlocal schemes showed less bias than the MYJ (local), which had insufficient BL mixing. They later looked at an improved YSU scheme (Hu et al. 2013) and concluded that the improvements appear to have corrected the overestimates in the temperature and wind profiles. The schemes under-predict temperature and over-predict moisture, and at nighttime the YSU scheme produces higher temperatures and winds. The ACM2 profiles appeared slightly better than the other schemes in this study.

Wang et al (2011) performed a spectral analysis of wind speed time series on the results from 4 BL schemes and found that there were no pronounced differences between them. However, the simulation results from the QNSE are the closest to observation for their simulations in the arid climates, scarce vegetation, and complex terrain of China.

Shin and Hong (2011) evaluated the YSU and ACM2 schemes (first-order closure) and 3 TKE schemes (MYJ, QNSE, and BouLac) for CASES-99 over Kansas using WRF v.3.2 and found that all schemes have difficulty at night, having large biases. Under statically stable conditions, no scheme performed satisfactorily. Neither the surface BL nor the upper inversion was resolved well, but the TKE schemes generally outperformed the first-order schemes.

Balzarani et al. (2014) analyzed the performance of 5 PBL schemes—ACM2, MRF, YSU, MYJ and UW—using a 5-km grid spacing domain of WRF v.3.2 and found that all 5 parameterizations produce similar performances in terms of temperature, mixing ratio, and wind speed near Milan, Italy, implying some systematic errors in all simulations. However, their analysis was limited only to model bias with major application to air quality. Their study concluded that temperature forecasts seem to impact the early evolution of the PBL depth, while entrainment fluxes parameterizations have major influence on the afternoon development. Over-prediction of temperature and wind speed produce excessive mixing. The MRF, MYJ, and ACM2 schemes use a similar approach in representing the entrainment process and overestimate the PBL depth. The best agreement between model and both Lidar and balloon observations was identified as YSU.

HariPrasad et al. (2014) compared the MYJ, MYNN, QNSE, YSU, ACM2, BouLac, and UW schemes using sonde, met tower, and sonic anemometer data over India. For WRF v3.4 they found that MYNN and YSU were in closer agreement with the observations. Shrivastava et al. (2014) examined the BL/SL combinations of YSU-Monin-Obukhov (MO), ACM2-MO, MJY-MO, and QNSE-QNSE BL/SL scheme pairings on a 3-km grid over India and found that the MYJ-MO scheme pairing performed slightly better than the others tested. In addition, the

Development Testbed Center (DTC) compared 2 configurations of the then AFWA (now 557th Weather Wing) operational schemes, ACM2-Pleim-Xiu (PX), and YSU-PX BL/SL configurations and the relative performance was related to time of day (DTC 2015).

Cuchiara et al. (2014) tested the YSU, ACM2, MYJ, and QNSE schemes over a Houston, Texas-centered domain and obtained high correlation coefficients and small biases for all of the meteorological variables tested. Variables analyzed included the temperature, potential temperature, mixing ratio, relative humidity (RH) and the u and v wind components. Despite similar performance, the YSU scheme provided the best agreement with ozone observations.

In a relevant recent study focusing on the prediction of near-surface and PBL properties over complex terrain, Dimitrova et al. (2014) examined 6 PBL options (YSU, MRF, ACM2, MYJ, BouLac, and QNSE). Extensive observational data were available from the MATERHORN program. While all over-predicted the minimum temperature inside the “valley cold pool”, the QNSE scheme was the best performer for near-surface temperature (2m) and wind (10m).

While our primary research interests are not in tropical storms or in air quality but in complex terrain environments for land mobility, these studies are useful in illustrating how the scheme selection process can be a function of the geography and climatology of the model domain. In summary, none of the earlier studies showed a clear preferred scheme that could be universally applied to all geographic regions, season, and time of day. The WRF model is not static, and physical schemes are being continually added, removed, and modified. Finally, we have the issue of verification. In most cases, observational data were used. However, where there are insufficient observations, sometimes LES models are used as a proxy (Ayotte et al. 1996). Therefore, it is challenging to find a single configuration that works optimally in all environments, and under all synoptic conditions and geographic regions.

3.3 Direction

What should become clear is that each scheme is designed using a particular set of observations, LES simulations, or BL conditions as validation. Validation comes from a variety of methods (e.g., observations and LES models) and different meteorological conditions. In addition, the physics of the stable and unstable PBL are quite different, leading to a “stability dependency” between the various schemes. While one particular scheme may perform best under certain conditions, another may be optimal under entirely different meteorological conditions or environments. A variety of local, nonlocal, and hybrid schemes are represented.

Since hybrid schemes have a local and nonlocal component, they have the potential for representing large- and small-scale eddies well. The purpose of this research is to investigate the utility of a single scheme in complex topography and under a reasonable variety of land use and flow regimes. Another objective is to examine the feasibility of using BL/SL parameterization combinations to construct a simple physics-based ensemble that would provide crude measures of forecast uncertainty or confidence.

4. Domain and Model Configuration

The numerical model used in this analysis, the WRF-ARW (Skamarock 2008), is a community weather forecast model designed and supported by NCAR. For the purposes of this study, WRF v3.8.1 was selected. This community model has a major update every year with minor revisions and bug fixes applied about 4 months later. At the time of this study, v3.8.1 is the latest version of the modeling system that includes the minor revisions; hence, it was selected. The WRF has an advantage over most current operational models such as the Global Air Land Weather Exploitation Model (GALWEM, Stoffler 2016) and the GFS because it includes a variety of user-selectable physical parameterization scheme options. This is advantageous in constructing physics-based ensembles that can yield a measure of model uncertainty due to the included model physics. Aside from the selection of various physics options, initial conditions may be varied easily from within the model construct.

The model specifications common to all 7 experiments using WRF-ARW as employed in this study are shown in Table 3.

Table 3 WRF-ARW common namelist options used for this study

Namelist parameter	Option selected
Shortwave radiation	Dudhia Scheme
Longwave radiation	RRTM
Explicit moist microphysics	Thompson
Cumulus parameterization	Kain-Fritsch 9 km only, explicit 1,3 km
PBL Scheme	Varies (see table 5)
Surface layer	Paired with PBL Scheme (Table 5)
Land Surface Scheme	NOAH
Time step to grid ratio (seconds:kilometers)	3:1
Horizontal subgrid diffusion	Second-order on coordinate surfaces
Subgrid turbulence closure	Horizontal Smagorinsky first-order

Table 3 WRF-ARW common namelist options used for this study (continued)

Namelist parameter	Option selected
Number of vertical terrain-following levels	57
Vertical velocity damping	Yes
Feedback (2-way interactive)	Yes
Nesting	Yes
Terrain slope shadow	Yes
FDDA	Yes
Nudging strength	$4.0 \times 10^{-4} \text{ s}^{-1}$

The FDDA option used in these model runs is based on observation nudging (Liu et al. 2005; Deng et al. 2009). This option is much less computationally expensive than traditional 4-D variational data assimilation (Huang et al. 2009) or ensemble Kalman filtering (Zupanski et al. 2008) and is thought to be a viable method for synoptic meteorological observation assimilation and deterministic fine-scale modeling applications such as those the Army is interested in. Schroeder et al. (2006) describes some earlier work that includes a nudging data assimilation component for Army applications. For a forward-deployed system, it is important that we incorporate computationally conservative techniques as much as possible.

Observation nudging is employed within the WRE-N framework because ARL is interested in integrating the continuous data assimilation strategy of observational nudging as an additional tool for assimilating forward battlefield observations not routinely ingested into the upstream boundary condition model (e.g., from coarser grid models provided by the 557th Weather Wing or other sources) that apply an intermittent or cycling 3-D variational data assimilation system for WRF or MM5 (Surmeier and Weigel 2004) or newer techniques within GALWEM (Stoffler 2016). These observations include, but are not limited to, local surface mesonets and sensors, local rawinsonde/dropsondes, sensors onboard aircraft, and unmanned aerial vehicle sensors. Tropospheric Airborne Meteorological Data Reporting (TAMDAR) measurements (Jacobs et al. 2014) were ingested separately. In these simulations, TAMDAR data were processed in addition to the Meteorological Assimilation Data Ingest System (MADIS) data for model initialization. The data sources chosen to be assimilated into the initialization for these runs are listed in Table 4.

Table 4 Initialization and nudging data switches (aside from coarse grid model data) used in the preprocessing package RUNWPSPLUS for WRF-ARW runs examined in this study

Setting name	Setting value
use_tamdar	Yes
use_madis_mesonet	Yes
use_madis_profiler_npn	Yes
use_madis_profiler_npn	Yes
use_madis_acars	Yes
use_madis_maritime	Yes
use_madis_metar	Yes
use_madis_raob	Yes
use_madis_sao	Yes
use_madis_satwind	No
use_madis_satwind1h	No
geog_data_res	2m, 30 s, 30 s

The steps involved in running the WRF model are shown in Fig. 5. The model is initialized using coarse grid model output from the GFS model. Additional inputs include 1/12° Real Time Global Sea Surface Temperature (Gemmil et al. 2007) sea surface temperature data (since part of the model domain is over water), 1-km snow data from the National Weather Service’s National Operational Hydrologic Remote Sensing Center Snow Data Assimilation System (NSIDC 2004), MADIS (NOAA, 2016) observations including surface, rawinsonde, Aircraft Communications Addressing and Reporting System, profiler, ship, and buoy data, as well as the TAMDAR data. The WRF preprocessing system ingests the model, sea surface temperature and snow data to construct initial and boundary conditions for the model. Lists provided by NCEP are used to determine MADIS mesonet observations that should not be processed due to a history of providing low-quality observations. The MADIS and TAMDAR data are then input to the Obsgrid program, which provides quality control to the observations and merges the GFS data with the quality controlled observations to provide initial conditions for WRF.

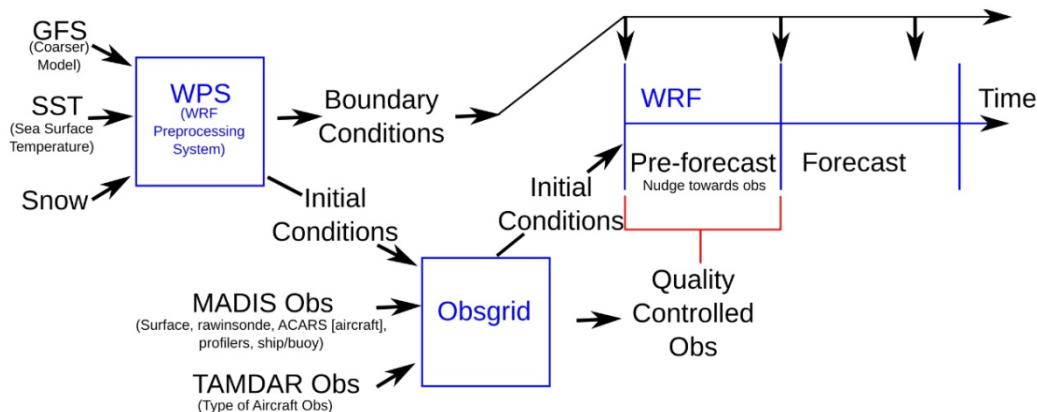


Fig. 5 Steps in the running of WRF-ARW at ARL, showing the ingest of meteorological and coarse grid data, preprocessing, and production of a WRF forecast preceded by a preforecast with observational nudging. In this study, the preforecast was 6 h and the forecast 18 h.

These steps are configured locally through a RUNWPSPLUS program (Kirby et al. 2013), which was written in Practical Extraction Report Language (Perl) scripting language by Dr Brian Reen of ARL. The RUNWPSPLUS program automates and configures much of the preprocessing for WRF. The program also downloads the specified (user-selectable) sources of initialization meteorological data for ingest into the modeling system except for the coarse grid model data (here, GFS). The boundary conditions and initial conditions generated via application of RUNWPSPLUS then provide input to the WRF modeling system. The program `real.exe` is used to prepare the input data for use by WRF, and `wrf.exe` is then run to create the forecast. Additionally, the first 6 h of the model run ingests data that nudge the forecast toward observations, so the 12Z run incorporates new observations via observational nudging until 18Z, and the forecast is computed after 18Z without additional observations. However, the nudging is gradually decreased between 18 and 19 coordinated universal time (UTC) with no new observations being assimilated, though the observation nudging is ongoing. Postprocessing is achieved using the Unified Postprocessor (UPP)* and the NCAR-supplied Model Evaluation Tools (MET) software (NCAR 2013), the latter developed at NCAR through grants from AFWA and the National Oceanic and Atmospheric Administration (NOAA). NCAR is sponsored by the United States National Science Foundation. Dawson et al. (2016) discusses automation of the UPP on a local workstation. The UPP ingests the NetCDF history files of the WRF to interpolate the model output to the National Weather Service standard pressure and

* The UPP is described in the user guide at: https://dtcenter.org/upp/users/docs/user_guide/V3/upp_users_guide.pdf.v3.0.

height levels and then generates output on model native levels as well as pressure surfaces. It computes the diagnostic quantities for several variables and generates all output on an unstaggered Arakawa A Grid in hourly gridded binary (GRIB) files, Edition 1. The UPP GRIB files are compatible with MET software.

This study uses a triple nest configuration of WRF-ARW centered approximately near San Diego, California. The domain includes a variety of geography, land use, and topography. The marine environment, coastal environment, large valley, gently sloping desert, and mountainous terrain are all represented. Regarding land use, there are urban, suburban, agricultural, grassland, arid, and mountainous areas with a variety of subtropical and mid-latitude vegetation including grassland and forest. The domains are shown in Fig. 6. WRE-N was employed an outer 9-km nest of 175×175 grid points (1566×1566 km) (D1), 242×241 grid points on the middle 3-km nest (720×720 km) (D2), and 127×127 grid points on the 1-km inner nest (126×126 km) (D3). The model top was selected to be 10mb, which is quite conservative.

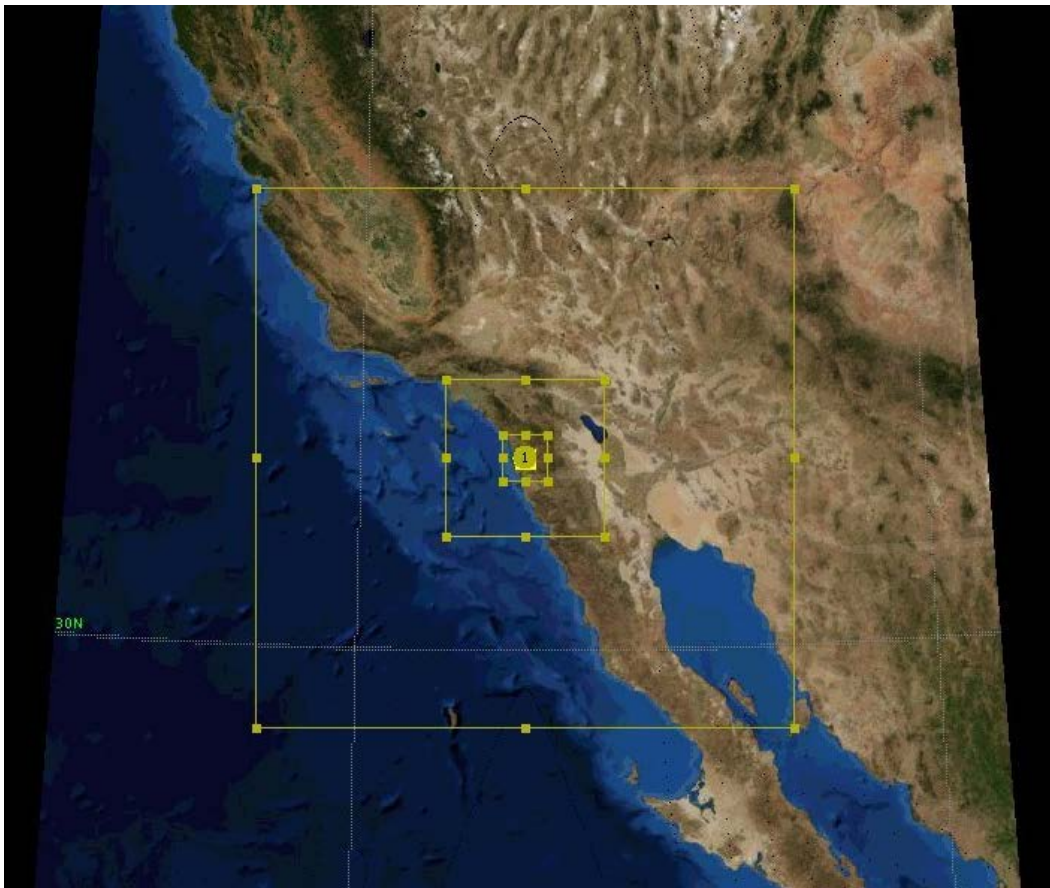


Fig. 6 Map of the model domain for these experiments. The configuration consists of a triple nested domain (9-3-1 km) centered approximately over San Diego, CA. It is the same domain used for the ARL GIS studies (Smith et al. 2015).

The southern California dataset was used for this study. There were 5 days of intensive observations available for data initialization and validation (Table 5). Four of the days were characterized by active weather events over the model domain; one day was characterized as quiescent, with a large ridge centered over the forecast area. For the purposes of this study, only data during quiescent conditions (9–10 February 2012) were considered. Additional simulations were also run for the other 4 days. We chose to concentrate on this case since the primary intent of this study was to examine the model’s ability to capture the development of the daytime CBL and transition to the nocturnal BL. Examining the CBL development under quiescent conditions minimizes the effects of larger synoptic-scale weather features that can complicate and suddenly modulate the development of the daytime CBL. In the absence of large-scale forcing (except for subsidence), cloudiness, advection, and precipitation, the processes controlling the depth of the CBL and its growth should be highly simplified.

Table 5 Synoptic conditions for the case study days considered

Case	Dates (2012)	Description
1	February 7–8	Upper-level front moved onshore, which led to widespread precipitation in the region.
2	February 9–10	Quiescent weather was in place with a 500-hPa ridge centered over central California at 12 UTC.
3	February 16–17	An upper-level low located near the California/Arizona border with Mexico at 12 UTC brought precipitation to that portion of the domain. This pattern moved south and east over the course of the day.
4	March 1–2	A weak shortwave trough resulted in precipitation in northern California at the beginning of the period that spread to Nevada, then moved southward and decreased in coverage.
5	March 5–6	Widespread high-level cloudiness due to weak upper-level low pressure but very limited precipitation.

Case 2 was used for all of the model runs used in this analysis. The dependency upon synoptic conditions is examined in the ongoing Design of Experiments (DoE) study (Smith and Penc 2015a, 2015b; Smith et al. 2016a, 2016b; Smith et al. 2017, 2018). The DoE study examines data from all 5 days of intensive data collection over 2 domains that overlap: the current study domain (centered over San Diego [KSAN]) and an adjacent domain centered over San Francisco, California (KSFO). The surface and 500-hPa maps covering the study area are shown in Figs. 7 and 8. The 500-hPa charts for 1200 UTC 09 February and 0000Z 10 February are shown in Figs. 7a (left) and 7b (right), respectively, and show ridging in place over the

study area, with the 500-hPa ridge axis slightly to the north of the innermost mesh, which is located near KSAN.

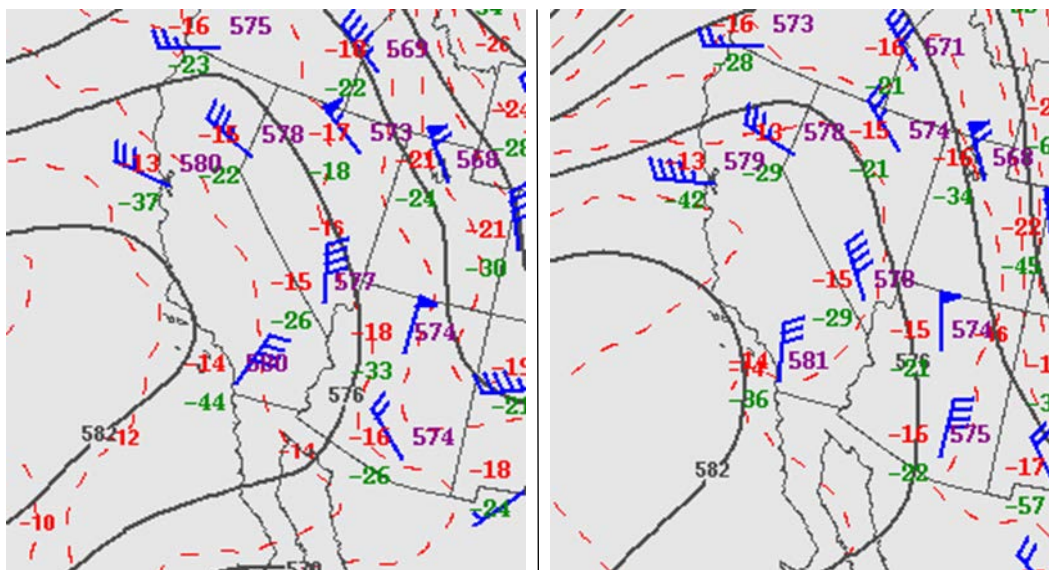


Fig. 7 Analysis (500 hPa) for 1200Z 09 February 2012 (left) and 0000Z 10 February 2012 (right). (Reprinted from NOAA, <http://www.spc.noaa.gov/obswx/maps/>)

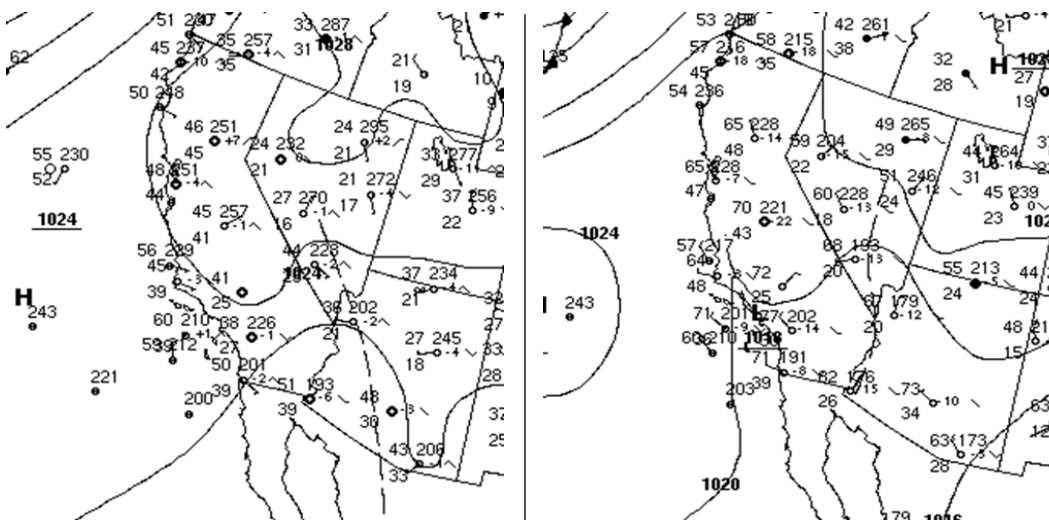


Fig. 8 Surface weather maps for 1200Z 09 February 2012 (left) and 0000Z 10 February 2012 (right) (Reprinted from NOAA, http://www.wpc.ncep.noaa.gov/archives/web_pages/sfc/sfc_archive.php)

The corresponding surface charts are shown in Figs. 8. The surface high-pressure center remained to the west of California during the study period, with generally light surface winds observed. Cloud coverage was light over the study area.

Synoptic analyses indicate that a light flow regime at the surface was in place during the study period. A time series of the observed surface wind speeds at North Island Naval Air Station, San Diego (KNZY; latitude 32.69°N, longitude 117.21°W, and elevation 23 ft) is shown in Fig. 9. Prior to 2100 UTC, the surface winds are uniformly light at approximately 2ms^{-1} during the early morning through local noon. Winds pick up during the afternoon with the development of surface heating and the daytime CBL. After sunset, winds again decrease to near 2ms^{-1} . For the majority of the simulation, the winds remain light. For the time period that the nocturnal BL is in place, winds are nearly constant and light and in the vicinity of 2ms^{-1} .

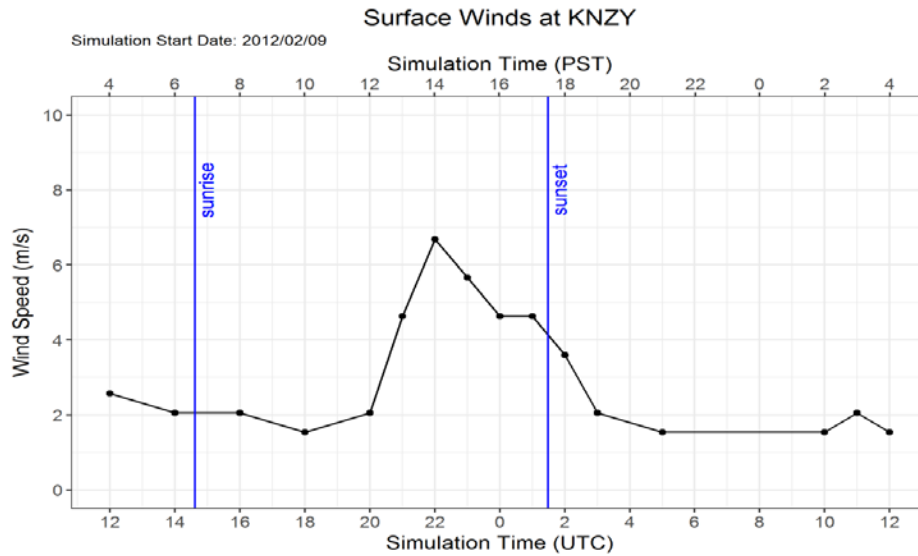


Fig. 9 Time series plot of the surface (2m) winds at KNZY; latitude 32.69°N, longitude 117.21°W, and elevation 23 ft

With quiescent conditions at play during the study period, the focus of this analysis was to examine the representation of the BL by the candidate schemes, with emphasis on capturing surface conditions during the diurnal evolution of the PBL. The results of this study follow in Section 5.

5. Results

The 7 PBL/SL parameterization combinations selected for this study are listed in Table 6. In each of these, the default SL scheme was coupled with the PBL scheme being tested because they are generally accepted and the most widely used by researchers. Where there was a matching SL, we used that option. For the others, where multiple options existed for SL choice, we used the revised MM5 SL scheme. That scheme is reported to work with many of the PBL options. We did not deviate

from the suggested configuration to limit the total number of runs, as this is anticipated to be the best performing combination.

Table 6 PBL/SL combinations tested and evaluated in this study

Case no.	PBL/SL option	PBL scheme	SL scheme
1	5 / 5	Mellor-Yamada Nakanishi and Niino (MYNN)	Nakanishi and Niino PBL's SL scheme
2	11 / 1	Shin-Hong Scheme (SH)	Revised MM5 SL scheme
3	2 / 2	Mellor-Yamada-Janjic Scheme (MYJ)	Eta similarity SL scheme
4	1 / 1	Yonsei University Scheme (YSU)	Revised MM5 SL scheme
5	8 / 1	Bougeault-Lacarrère PBL (BouLac)	Revised MM5 SL scheme
6	4 / 4	Quasi-normal Scale Elimination (QNSE)	QNSE PBL scheme's SL option
7	7 / 1	Asymmetric Convective Model (ACM2)	Revised MM5 SL scheme

Model performance was assessed using point observations and a statistical analysis and comparison of the model bias and root-mean-squared error (RMSE) for the 7 PBL/SL scheme combinations tested. In addition, a number of other statistical measures were derived. The 6 meteorological variables analyzed in this analysis were temperature (T), dew point (DPT), relative humidity (RH), u component of the wind (u), v component of the wind (v), and wind speed. We also briefly examined the PBL depth, as deduced from the model output, and performed an analysis of the ensemble spread, which we present in the form of Talagrand diagrams (Hamill 2001).

5.1 Model Bias and RMSE

Model bias and RMSE were calculated for each of the 7 members in the ensemble. Statistics were calculated for each hour of the simulation, including the data assimilation (hours 1–6), the nowcast period (hours 7–12), and the extended forecast (hours 13–24). Over the model domain, these periods roughly correspond to morning, afternoon, and nighttime hours, respectively. This corresponds to, respectively, 1300 UTC 09 February 2012 through 1800 UTC 09 February (0500–1000 Pacific Standard Time [PST]), 1900 UTC 09 February 2012 through 0000 UTC 10 February 2012 (1100–1600 PST), and 0100 UTC 10 February 2012 through 1200 UTC 10 February 2012 (1700–0400 PST). In addition, the overall

model statistics were computed (hours 1–24) and reflect the period 1300 UTC 09 February 2012 through 1200 UTC 10 February 2012.

We calculated the bias by taking

$$B_i = F_i - O_i, \quad (1)$$

where F and O represent the forecast and observed value, and i indicated the specific pair in a given run.

We calculated the mean error or mean bias via

$$\text{mean error} = \frac{1}{N} \sum_{i=1}^N B_i \quad (2)$$

and the root mean square error via

$$\text{root mean square error} = \sqrt{\sum_{i=1}^N B_i^2}, \quad (3)$$

where N is the number of forecast observation pairs in a given run.

All statistics were computed using the R statistical language (R Core Team 2013) after compiling and postprocessing all the run data using the Tidyverse set of packages (Wickham 2011, 2014, 2017a). We plotted the various statistics and parameters of interest using the R statistical language and specifically the ggplot2 plot package (Wickham 2009, 2016, 2017b).

5.1.1 Model Bias

Model bias throughout the model integration is shown in Figs. 10–15. Each plot is labeled in terms of both UTC and PST with indicators for the times of sunrise and sunset. All of the analyses presented here apply to the innermost D3 (1-km) domain. The temperature bias is shown in Fig. 10. During the data assimilation period (1200–1800 UTC) all of the schemes show a small forecast bias, typically less than 1 K. While the ACM2, BouLac, MYJ, SH, and YSU schemes perform very similarly, the MYNN and QNSE schemes differ from these. The latter 2 schemes exhibit a negative bias, meaning that they underestimate the surface temperature. The former schemes slightly overestimate the surface temperature at 2m for the hours 1400 through 1600 UTC. After hour 4 of the simulation, all of the schemes converge and tend toward underestimating the surface temperature.

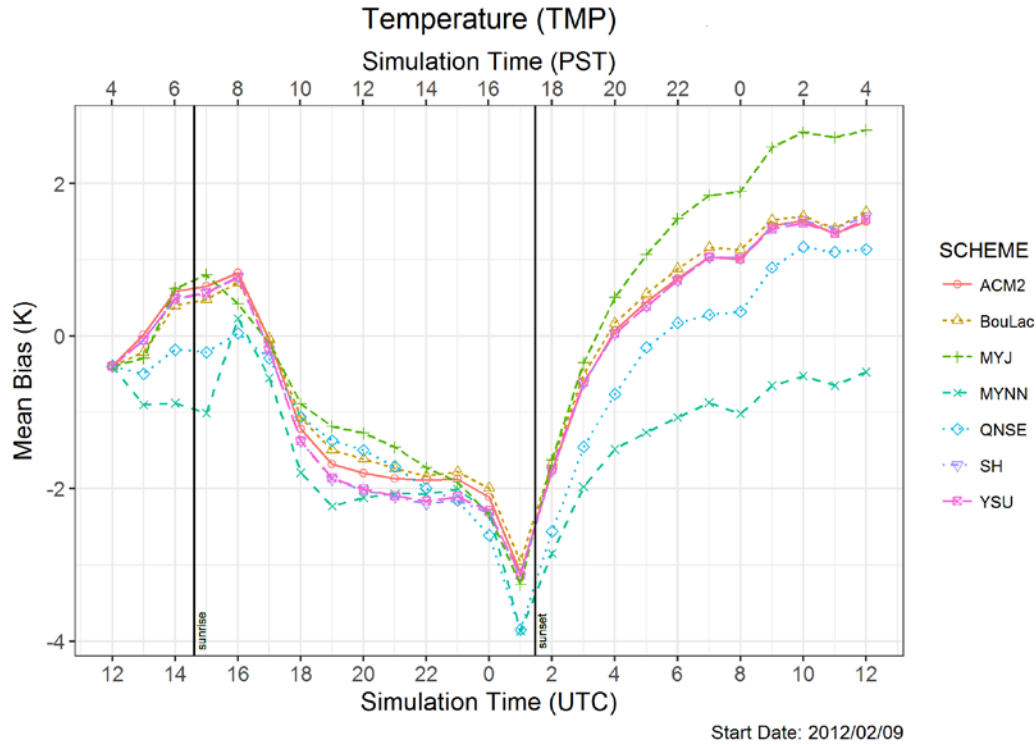


Fig. 10 Model bias for surface temperature (Kelvin [K]) at 2m above ground level (AGL) for the 7 PBL schemes tested. The data assimilation was applied from 1200 to 1800 UTC, and thus 0000 UTC is the beginning of the nowcast.

Until hour 12 of the simulation (0000 UTC), the model consistently underestimates the 2m surface temperature, and the negative temperature bias increases until approximately 0100 UTC on 10 February, when the negative bias is largest. The difference between the schemes is relatively small throughout the time period and amounts to only a couple tenths of a degree. Beyond 0100 UTC, the model bias reverses and overall becomes slightly positive (2m temperature is overestimated) during the final 4 h of the simulation, with the exception of the MYNN scheme, which retains a negative bias. This time period corresponds to nighttime, when a nocturnal BL is in place.

The MYJ scheme has the largest positive bias in the latter portion of the forecast period, which occurs during nighttime. The bias exceeds 2 K for the last 4 h of the simulation. Since the SH scheme is based largely on the YSU scheme, they perform very closely throughout the forecast period, and the curves largely superimpose. Based on the data, we conclude that the change in behavior is likely associated with the transition from the daytime CBL to the nighttime and the model's ability to effectively simulate both. During the daytime, the model underestimates the surface temperature, and at nighttime it overestimates the surface temperature, in general,

regardless of the scheme chosen, with the sole exception being the QNSE scheme, which retains a negative bias throughout the forecast cycle.

The bias for the DPT is shown in Fig. 11. For the majority of the forecast cycle, the DPT (at the surface) is underestimated by the model as shown by the negative bias, with the exception being between 0000 and 0200 UTC (1600–1800 PST). During these 3 h, which occur in late afternoon leading into early evening, the DPT is overestimated. Throughout the simulation, there is considerably more spread among the individual schemes than was seen in the temperature bias.

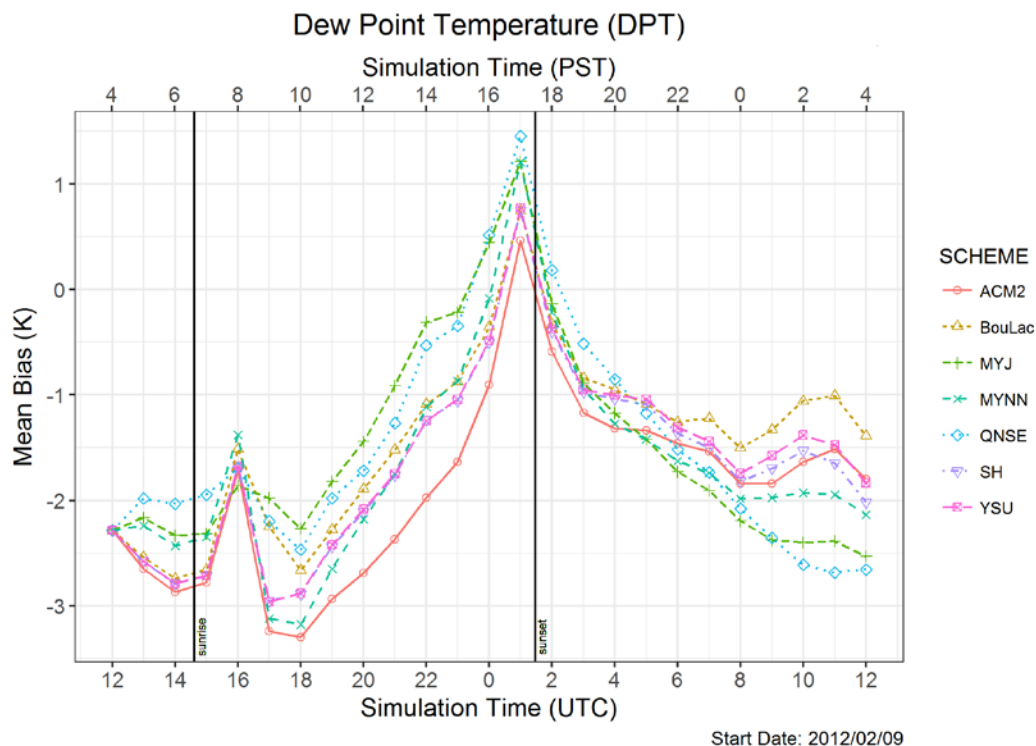


Fig. 11 Model bias for DPT (K) at 2m AGL for the 7 PBL schemes tested. The data assimilation was applied from 1200 to 1800 UTC, and thus 0000 UTC is the beginning of the nowcast.

The reversal in the bias tendency (change of bias versus time) corresponds to late afternoon, when the CBL is beginning to decay and be replaced by the nocturnal BL and corresponds closely with local sunset. At nighttime, during the latter half of the simulation, the forecast DPT is again underestimated, with overall negative bias being shown by all of the schemes tested. All of the schemes show a large range in the bias (defined by subtracting the maximum and minimum biases for each scheme) throughout the simulation—nearly 4 K. These changes in the bias tendency occur when the BL transitions between daytime and nighttime regimes. As in the previous analysis of surface temperature, there appears to be a dependency between the time of day and the bias as well as the portion of the forecast cycle

(assimilation, nowcast, and extended forecast). Since the model is initiated at 1200 UTC (0400 UTC) and it takes from 3 to 4 h (in our cases) for the model to stabilize and spin up, the transition between the stable nighttime BL and daytime CBL is not as clearly defined. The change in bias trend during this transition is likely a function of the drying and moistening of the PBL, which is controlled by BL processes, including entrainment and mixing into the PBL.

The RH bias for the simulations is shown in Fig. 12. The bias is largely similar to that of the DPT bias shown earlier. While during data assimilation there appears to be some spread among the schemes, with QNSE and MYNN performing slightly better, all members consistently show a negative bias. The initial bias is approximately 7%. The RH bias becomes steadily less negative almost from the beginning of the model simulation and reaches zero approximately 9 h into the forecast (2100 UTC, 1300 PST). The RH bias then becomes increasingly positive through much of the remainder of the afternoon (over-forecasting the RH), with a maximum positive bias of 8%–10% at 0100 UTC (1700 PST), which is just before sunset over the model domain. Thus, the model becomes too moist (in terms of RH), independent of BL scheme chosen, by late afternoon over the inner domain.

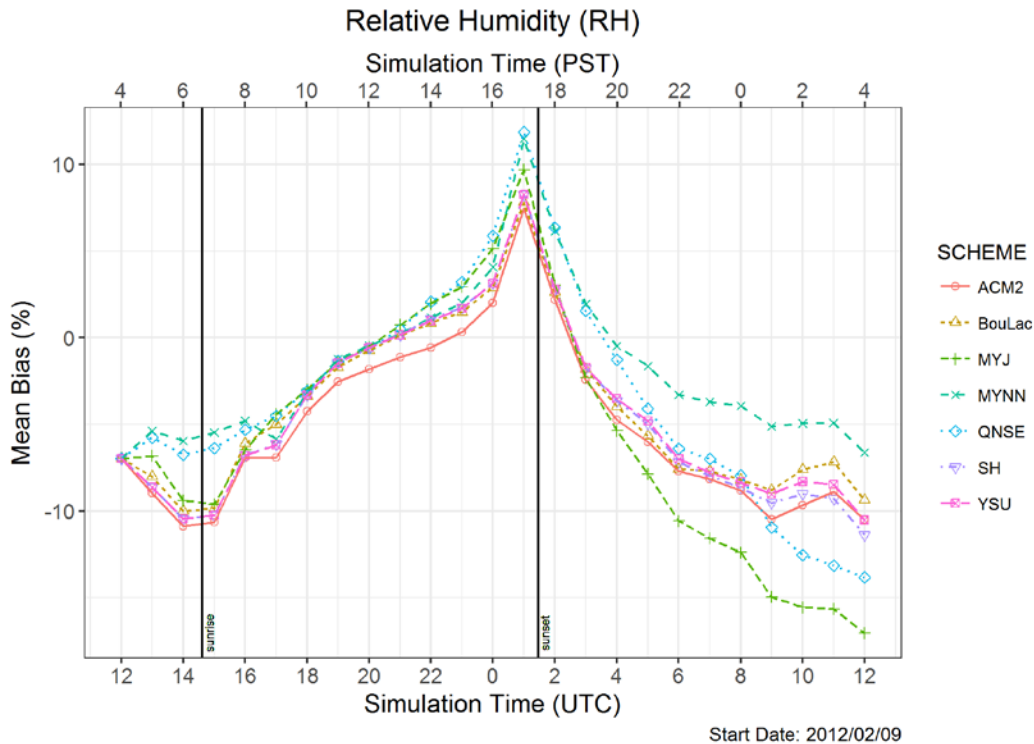


Fig. 12 Model bias for RH (%) at 2m AGL for the 7 PBL schemes tested. The data assimilation was applied from 1200 to 1800 UTC, and thus 0000 UTC is the beginning of the nowcast.

Immediately after model initialization, the bias for 5 of the schemes (ACM2, YSU, SH, BouLac, and MYJ) approaches -10% , while the other 2 schemes (MYNN and QNSE) approach a -5% bias, initially performing slightly better than the former schemes. After 1600 UTC (0800 PST) all 7 schemes performed approximately the same, with the model bias gradually and linearly approaching 0% by 2200 UTC (1400 PST), then the RH bias consistently increases to around 10% by 1700 UTC (1700 PST). While the model originally shows a negative bias, all 7 schemes are positively biased by mid-run. Thus, the model, regardless of scheme, produces surface (2m) RH that is consistently too high in late afternoon and early evening. After 0100 UTC, or 1700 PST locally, the model bias uniformly decreases throughout the remainder of the simulation. After 0400 UTC (2000 UTC) the model bias again becomes negative, corresponding to the onset of evening and nighttime over the model domain. As in the temperature analysis, the various schemes begin to diverge toward the end of the simulation (nighttime). The divergence in bias post-0400 UTC (2000 UTC) shows the MYNN scheme to be the best performer at approximately -9% RH bias, with the extremes being -14% bias (QNSE) and -18% bias (MYJ). The other schemes (ACM2, SH, BouLac, and YSU) are middle performers averaging close to -10% RH bias toward the end of the simulation. The model-produced RH is consistently too low at nighttime over the model domain. Interestingly, the peak in positive model bias around 0100 UTC corresponds closely to the peak in the negative temperature bias (Fig. 10). Since RH is a function of both temperature and moisture, this is not surprising.

The spread among ensemble members in the RH bias is largest toward the end of the simulation, which occurs in the early morning hours. This enhancement in the spread among the members is noted in all of the preceding bias analyses.

Figures 13–16 show the bias for the surface wind (at 10m AGL) data. The analyses are broken down into the east–west (u) component, north–south (v) component, and the wind speed. The individual components of the wind can be used to derive both a speed and direction, both of which can be in error. Errors in one or both of the individual components can be translated to directional errors, while the speed variable (WIND) translates to errors in the wind velocity. While the wind speed may be correct, the individual components may both be in error yet contribute to an apparent “correct” speed. In this case, the direction is in error but not in speed.

Figure 13 shows the bias for the u component of the wind. The mean bias for all 7 tested schemes does not differ significantly during the data assimilation part of the simulations, extending through about 0900 PST (1700 UTC). The initial mean bias of -1 ms^{-1} goes from negative to positive within 4 h of model initialization, then the ensemble mean approaches zero by the middle period of the simulation. The QNSE and MYJ schemes overshoot the ensemble mean between 1800 and 0000 UTC, then

settle with the remainder of the members through the end of the simulation. After the end of data assimilation (i.e., after 1800 UTC), during the nowcast period, the bias associated with each of the schemes varies the most. The bias from the other 5 schemes tested did not differ significantly from each other and averaged a near zero forecast bias. Past 0000 UTC (1600 PST), the bias values from all of the schemes behaved quite differently. The smallest bias error was associated with the MYNN scheme, followed by the ACM2 and MYJ schemes.

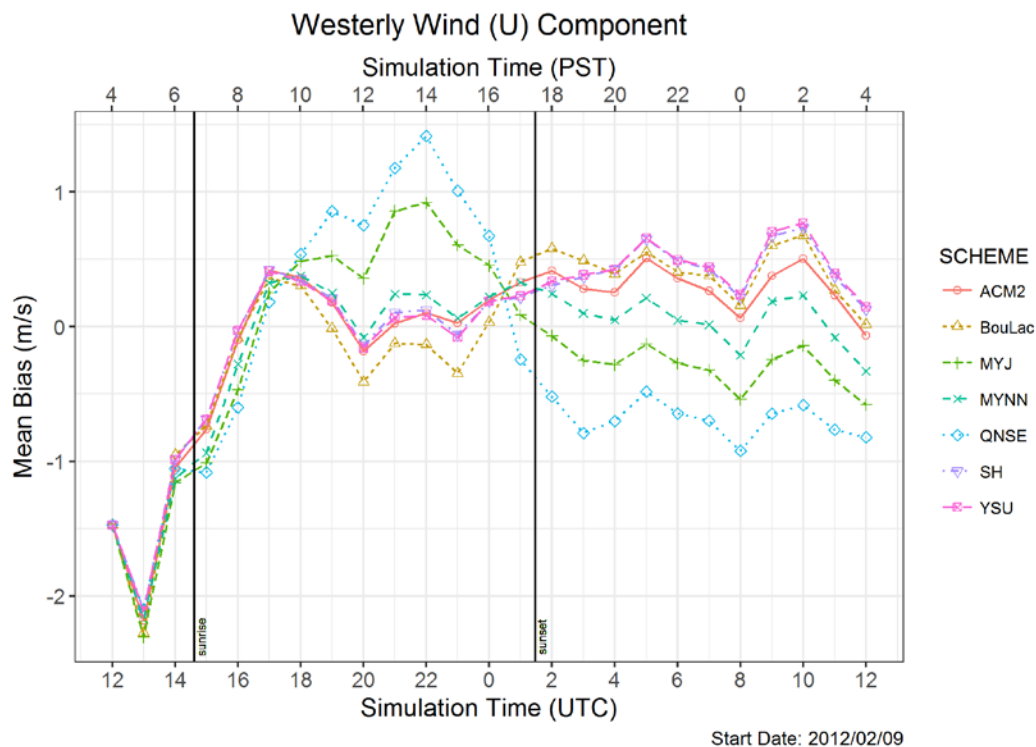


Fig. 13 Model bias for u component of wind (ms^{-1}) at 10m AGL for the 7 PBL schemes tested. The data assimilation was applied from 1200 to 1800 UTC, and thus 0000 UTC is the beginning of the nowcast.

As in the previous analyses, the bias results varied with the stage of the model run. Again, note that the extended forecast represents the daytime period over the study domain. The largest temporal change in the bias occurred during the model data assimilation phase, when the model was being adjusted to the observations. After this period, the model bias was steadier, with a shift around sunrise. The largest underestimate of the u component of the wind occurred right after model initialization, around 2 ms^{-1} . The relatively small magnitude of the bias error at later times must be viewed in context of the light surface winds and the ridging present over the model domain. Interestingly, the mean bias of the ensemble for the nowcast and the extended forecast is approximately zero.

The v component bias is shown in Fig. 14. As in the previous analysis, there is little difference between the individual PBL schemes during the earlier part of the simulation. The largest deviation between the schemes occurs during the latter half of the simulations. The better schemes in this case appear to be YSU, SH, and MYJ, which present the smallest biases of the 7 members analyzed. The worst performers are ACM2 and QNSE, at least when the v component of wind is considered. All of the schemes did give relatively small biases of less than 0.5 ms^{-1} , but keep in mind that the surface winds were light during most of this period. The largest model biases occurred during data assimilation (bias underestimates $>0.5 \text{ ms}^{-1}$) and the earlier part of the nowcast period, with bias errors approaching $0.7\text{--}1 \text{ ms}^{-1}$ around 8 h into the simulation. The latter time period corresponds to afternoon over the model domain. Except for the preforecast and nowcast periods near the beginning of the runs, the model ensemble does not appear to have a strong positive or negative bias. The mean bias of the ensemble is approximately zero for the latter half of the simulation time.

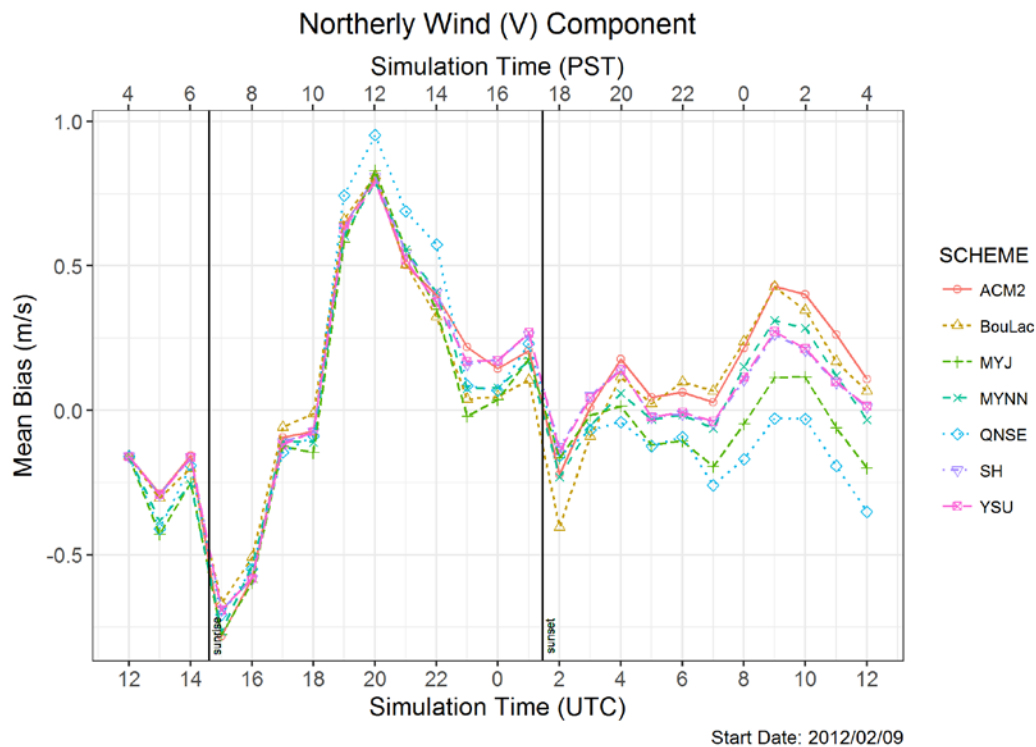


Fig. 14 Model bias for wind u component (ms^{-1}) at 10m AGL for the 7 PBL schemes tested. The data assimilation was applied from 1200 to 1800 UTC, and thus 0000 UTC is the beginning of the nowcast.

When we consider the surface wind speed (Fig. 15) bias we again see the different behavior in the bias for the 3 periods: assimilation, nowcast, and extended forecast. The model overestimates the surface wind during data assimilation by as much as

2 ms^{-1} but then quickly settles down to near zero bias by 4 h into the simulation. There is more spread in the model bias by the time the nowcast period is approaching, with the greatest bias occurring with the QNSE scheme. QNSE overestimates the wind speed by $>1 \text{ ms}^{-1}$ from hours 9–10 of the simulation. In contrast, the ACM2 scheme underestimates the wind by as much as 1 ms^{-1} around 8 h into the simulation. As in all of the previous bias analyses, the model bias contrast between the 7 schemes is most pronounced for the latter half of the simulation time. For wind speed bias, the best performers at nighttime are YSU, SH, and ACM2, with near zero bias for the extended forecast period. The worst performer for the extended forecast is QNSE, followed by MYJ, with a mean bias error of approximately 1 ms^{-1} , meaning that the model with these schemes overestimated the surface wind by this amount. Again, we must keep in mind that the observed surface winds were light in this case, with surface ridging over the model domain. Note that the physics differences appear to be most pronounced during nighttime. Examination of the 6 meteorological variables is consistent in this regard, at least for this study period.

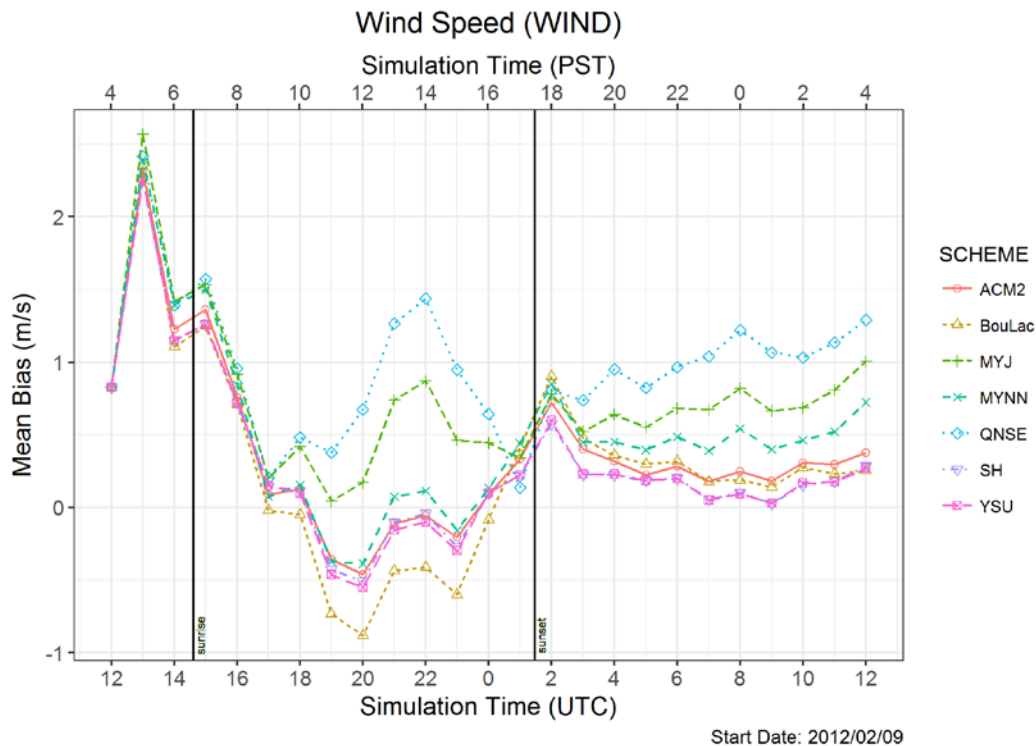


Fig. 15 Model bias for wind velocity (ms^{-1}) at 10m AGL for the 7 PBL schemes tested. The data assimilation was applied from 1200 to 1800 UTC, and thus 0000 UTC is the beginning of the nowcast.

In summary, these analyses show that consistently the model performance exhibits 3 distinct regimes. Those include data assimilation (0–6 h), the nowcast period

(6–12 h into the simulation), and the “extended” forecast. The differences in model bias between the schemes are very different in each regime. Initially, the various schemes show very little difference. During the nowcast period, after data assimilation is complete, there is considerable increase in the spread of the ensemble members. During the extended forecast period (12–24 h), the behavior is again quite different. The largest spread between members occurs during the latter part of the simulation. The best performing BL/SL combination is a function of the meteorological variable we select, the time of day, and the period we are examining (assimilation, nowcast, and extended forecast).

5.1.2 Model RMSE

Similar to the model bias analysis in the previous section, we also performed an analysis of the model RMSE. The surface (2m) temperature RMSE of each of the ensemble members is shown in Fig. 16. The RMSE of the members averages approximately 2.6 K at the beginning of the simulation time. The MYNN scheme shows the largest RMSE for the initial 4 h while QNSE shows the lowest. After 3 h of simulation time, the RMSE decreases to around 2 K for all of the members. For the nowcast period (6–12 h simulation time), the RMSE increases fairly linearly from about 1.5 to 3.5 K, with the highest RMSE associated with the MYNN and QNSE schemes. After 15 h, the RMSE of the members typically varies from 2.5 to 3.5 K. BouLac, ACM2, SH, and YSU follow each other closely toward the ending hours of the simulation in the extended forecast. Since SH is based on YSU, the similarity between these 2 schemes is not unexpected. The closeness of these schemes reflects the similarity in the formulation of the physics for the nighttime case, and these times reflect the nighttime scenario over the model domain. The increase of RMSE toward the end of the simulation is not unexpected since the errors tend to accumulate the further a forecast progresses in time from the point of initialization.

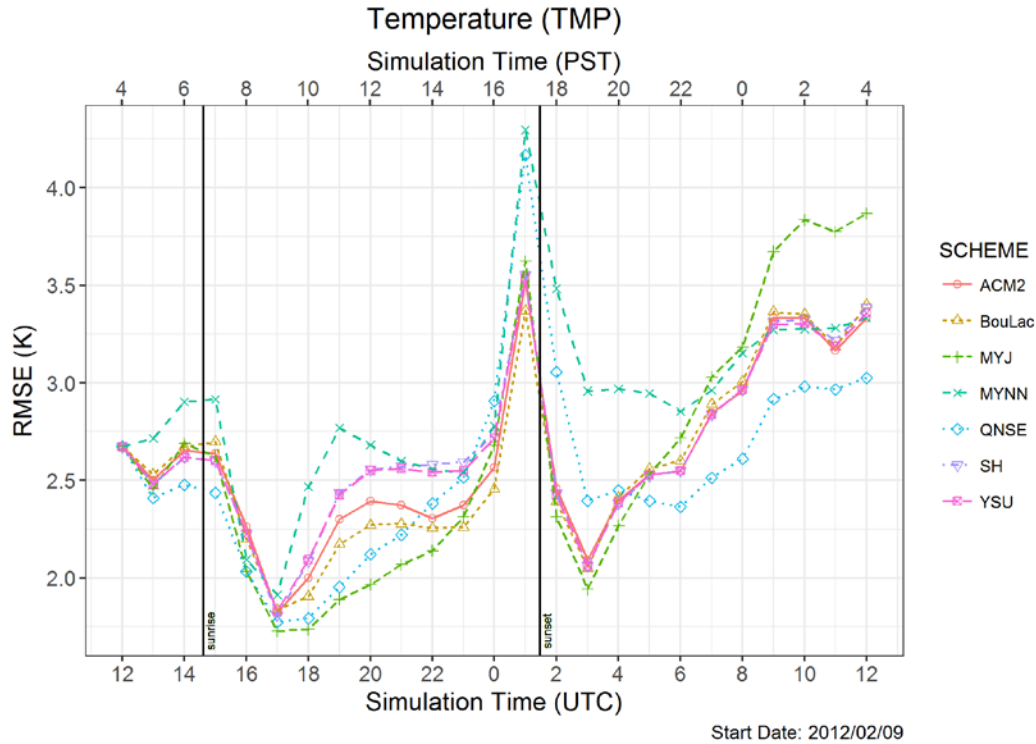


Fig. 16 Model RMSE for surface temperature (K) at 2m AGL for the 7 PBL schemes tested. The data assimilation was applied from 1200 to 1800 UTC, and thus 0000 UTC is the beginning of the nowcast.

Figure 17 shows the DPT RMSE for each of the schemes tested. The RMSE is rather large, reflecting model errors in the humidity and the temperature collectively. The mean RMSE at the point of initialization is approximately 3.3 K, which is significant. As data assimilation progresses, the RMSE increases to approximately 3.8 K, with an increased spread among the schemes tested. During the nowcast period (6–12 h into the simulations), the RMSE decreases from approximately 4.0 to 2.5 K. This corresponds to the time the daytime CBL is developing, and by late afternoon (1600–1800 PST) the RMSE is at the lowest point of the first half of the simulation. At the point in time that the nighttime BL is developing, early to late evening, the RMSE further decreases, with the mean RMSE falling to nearly 2.0 K averaged over all schemes. From midnight to early morning, the RMSE then increases for the model members from approximately 2.0 to 3.5 K. There is little spread among the model members, with no scheme performing notably better than any of the others. The most notable spread among the model members occurs during the early part of the nowcast period. MYJ, QNSE, and BouLac perform the best during this time with the lowest RMSE of all the members. MYNN and ACM2 have the highest RMSE of the 7 members during the period extending from the end of data assimilation through the early part of the

6-h nowcast period. With respect to the DPT RMSE, there is overall no clearly superior scheme to choose from.

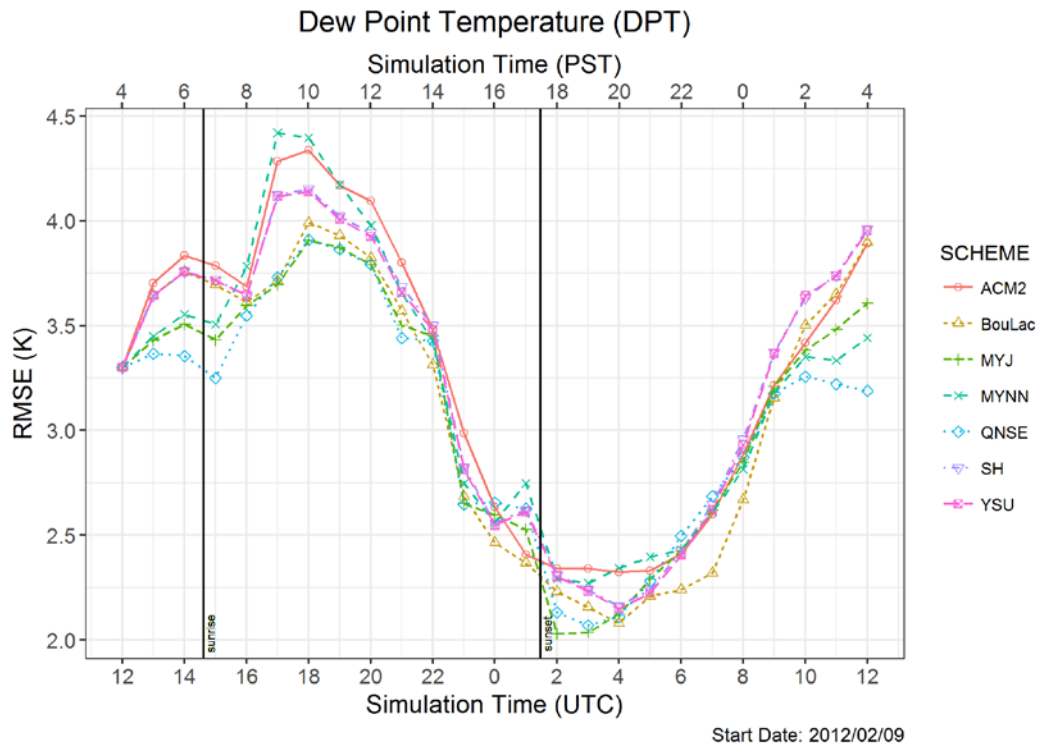


Fig. 17 Model RMSE for DPT (K) at 2m AGL for the 7 PBL schemes tested. The data assimilation was applied from 1200 to 1800 UTC, and thus 0000 UTC is the beginning of the nowcast.

Figure 18 shows the RH RMSE distribution for the 7 PBL/SL schemes we tested. After model initialization, 5 of the schemes (ACM2, YSU, SH, BouLac, and MYJ) edge toward approximately 17% RMSE, while the other 2 schemes (MYNN and QNSE) approach 15% RMSE and initially perform slightly better. Between the 4- and 6-h simulation time (data assimilation) all 7 schemes performed similarly. The model RMSE gradually and linearly approached 5%–7% RH RMSE, reaching a minimum RMSE by hours 10 to 11. Beginning at hour 11, the RMSE then begins a near monotonic increase to the end of the simulation (the notable exception being a spike in RMSE at 1700 PST) coincident with the timing of the collapse of the daytime PBL. Around simulation hour 9, the individual schemes begin to show some divergence in the RMSE with respect to the other schemes tested. At the onset of this latter period, the MYNN and QNSE schemes performed slightly better than the others; however, for the majority of the latter half of the simulation these 2 schemes were among the better performers. The MYJ scheme performed least well among the 7 toward the end of the simulation, with RMSE exceeding 20% RH. The MYNN scheme performed the best of the selections, with RMSE around 15% RH

toward the end when RMSE is maximum. In the middle of the pack, ACM2, SH, YSU, QNSE, and MYNN produced RMSE values that maximized at about 17% RH at the end of the simulation.

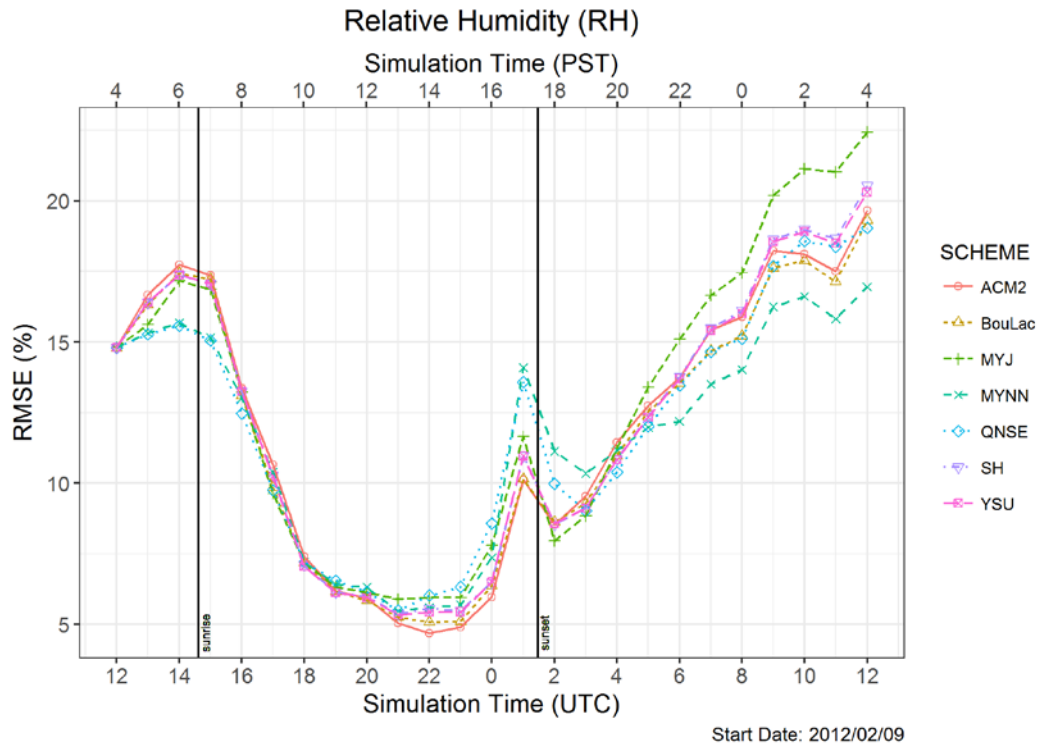


Fig. 18 Model RMSE for RH (%) at 2m AGL for the 7 PBL schemes tested. The data assimilation was applied from 1200 to 1800 UTC, and thus 0000 UTC is the beginning of the nowcast.

As in the previous analyses, the relative performance of each scheme, in terms of the percent-RH RMSE here, is dependent upon the phase of the forecast cycle and thus the time of day over the model domain. The RMSE decreases during the daytime part of the simulation and increases as nighttime progresses. This is consistent with our other findings.

Figures 19–21 show the RMSE for the horizontal wind components and the wind speed. The u component results are shown in Fig. 19. The largest RMSE occurs near the beginning of the simulation during data assimilation for all of the schemes. The RMSE approaches 3 ms^{-1} at hour 1 and then decreases rapidly for the remainder of the data assimilation period, reaching approximately 1.5 ms^{-1} by hour 5. The RMSE then diverges for the individual members, increasing for both QNSE and MYJ, and decreasing for the others (BouLac, ACM2, SH, YSU, and QNSE) by the midpoint of the simulation. At that point, the RMSE is approximately 1.5 ms^{-1} for each of the members. This time corresponds to late afternoon over the

domain. After the midpoint of the simulation, nighttime is settling over the domain, and the individual schemes show more spread. The RMSE for QNSE is highest, at approximately 1.9 ms^{-1} , followed by MYJ and MYNN. The remaining schemes were closely packed and had a mean RMSE of less than 1.5 ms^{-1} . Thus, the SH, YSU, BouLac, and ACM2 schemes performed slightly better for the nighttime portion of the simulation.

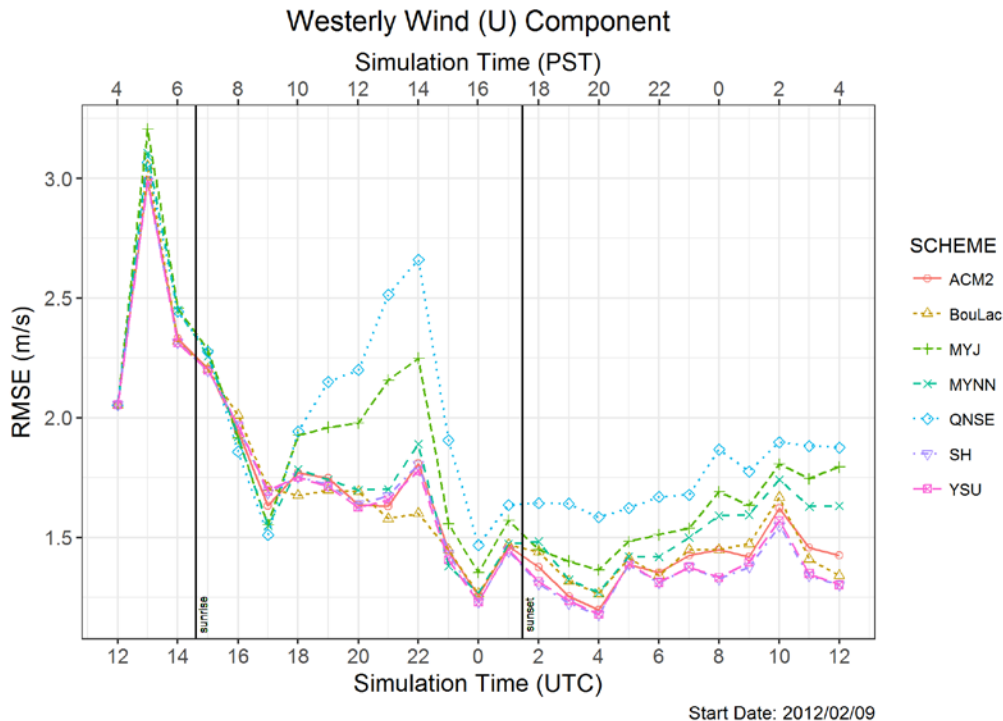


Fig. 19 Model RMSE for u component of wind (ms^{-1}) at 10m AGL for the 7 PBL schemes tested. The data assimilation was applied from 1200 to 1800 UTC, and thus 0000 UTC is the beginning of the nowcast.

The v component RMSE is shown in Fig. 20. All 7 schemes tested showed an immediate jump in RMSE at the beginning of the simulation, with the RMSE decreasing after hour 3 toward the end of data assimilation. BouLac performs slightly better than the others when RMSE is the measure. At hour 6, the end of data assimilation, the RMSE averages approximately 1.7 ms^{-1} for the members. The RMSE generally continues to decrease at this point in the simulation, with some divergence among members. From hours 6–12, QNSE shows a slightly higher RMSE than the others. BouLac, SH, and YSU perform the best in this regard. Up until hour 17, all of the schemes show a slight decrease in the RMSE. After 17-h simulation time, the RMSE remains approximately steady at approximately 1.4 to 1.5 ms^{-1} . There is a slight preference toward using the MYJ and SH schemes, in terms of lesser RMSE over the innermost model domain, while QNSE shows a slightly larger RMSE than the other members.

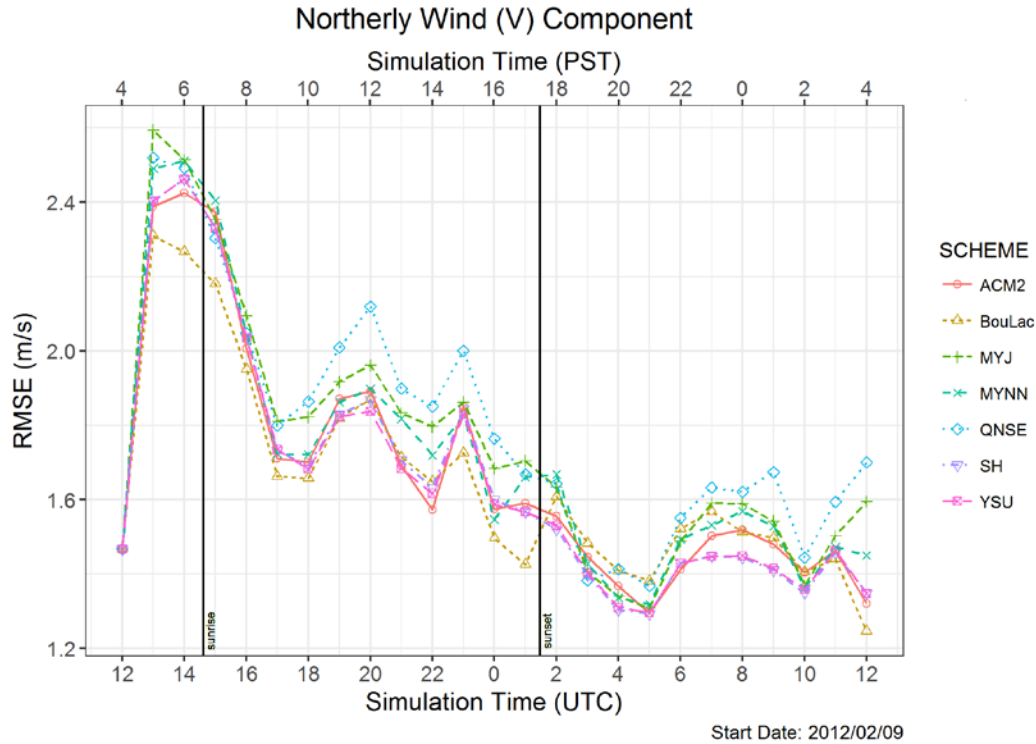


Fig. 20 Model RMSE for v component of wind (ms^{-1}) at 10m AGL for the 7 PBL schemes tested. The data assimilation was applied from 1200 to 1800 UTC, and thus 0000 UTC is the beginning of the nowcast.

The wind speed RMSE for the 7 schemes is shown in Fig. 21. At model initialization, the RMSE is 1.6 ms^{-1} . RMSE then increases to over 3 ms^{-1} in the first simulation hour, then settles to about 2.5 ms^{-1} the next hour. The RMSE continues to decrease for the next 5 h to around 1.5 ms^{-1} where it remains for the remainder of the model simulation time. There is little difference between the individual schemes during the first 5 h. After that, the spread between individual BL schemes increases. During the nowcast period, which corresponds to daytime, the QNSE scheme shows the greatest RMSE, followed by BouLac and MYJ. The remaining schemes (YSU, SH, YSU, and MYNN) have the lowest RMSE. The latter half of the simulation time, during the formation and maintenance of the nighttime BL, shows the greatest spread. MYJ, SH, BouLac, and ACM2, show the least RMSE, ranging from 1.2 to 1.5 ms^{-1} . The greatest variation between schemes occurs during the nighttime, when the nighttime-specific formulations are active. This again is consistent with the analyses of temperature and RH errors presented earlier. Differences between the various BL/SL formulations appear to be greater at nighttime rather than daytime.

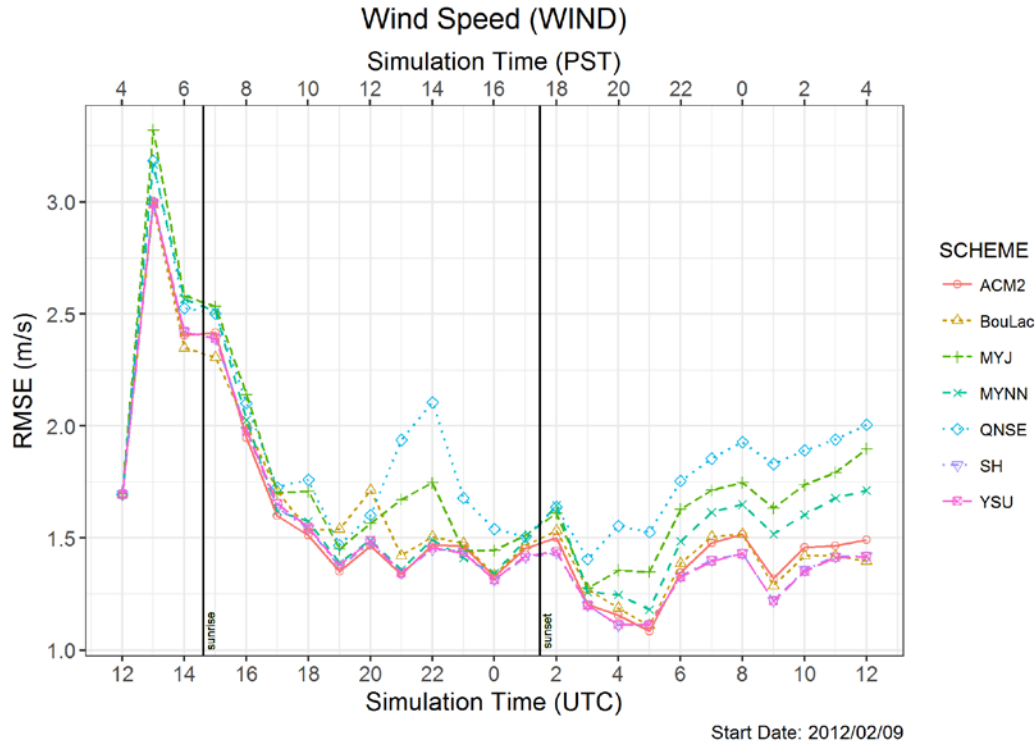


Fig. 21 Model RMSE for wind velocity (ms^{-1}) at 10m AGL for the 7 PBL schemes tested. The data assimilation was applied from 1200 to 1800 UTC, and thus 0000 UTC is the beginning of the nowcast.

As in the earlier figures, the relatively high errors are in association with model spin up, which takes approximately 3 h to achieve. At that point, the model error is quite stable, more or less predictable, and dependent not only on the scheme used, but on the time of day that the BL scheme is parameterized to represent. All of the analyses suggest that the largest differences between schemes occur within the nocturnal BL rather than the daytime CBL. This finding is consistent among all of the meteorological variables examined (T, DPT, RH, u component, v component, and wind speed).

5.1.3 Summary Statistics

Summary statistics of the mean bias and RMSE for the 6 meteorological variables were calculated for the BL/SL scheme combinations that we tested. Because the previous analyses showed that the performance of each scheme was highly correlated to the period of time in the simulation (i.e., data assimilation, nowcast, and extended forecast) and the time of day, these statistics were compiled for each of those time categories for comparison purposes. The data assimilation period coincides with the predawn to early morning transition, the nowcast to the development of the daytime CBL, and the extended forecast period to the decay of the daytime CBL and development of a nocturnal BL.

Table 7 shows the results of this analysis for the data assimilation phase of the forecast cycle, which runs from predawn to late morning. The best performers are colored red. The QNSE scheme is the best performer, as determined by the number of statistics (4) in which the scheme performed the best (bias and RMSE for each variable). The next best performers are the BouLac (3 statistics), SH, and ACM2 schemes (each with 2). In a modeling scenario, the best choice as to BL/SL scheme would be made depending on what variables are deemed to be to most important to the end user. Differences between each scheme are typically small, so there is a degree of flexibility in the selection process. There are no clearly outstanding performers in any of the categories we defined. Since this time period is the preforecast time period during which data assimilation is applied, the verification does not provide insight on the relative quality of the actual forecasts among the ensemble members. The preforecast time period corresponds to the latter few hours of the nocturnal BL and the earlier hours of the development of the daytime CBL.

Table 7 Mean bias and RMSE for 1-km ensemble members (D3) (1 km) for 1200 UTC 9 February through 1800 UTC 9 February (data assimilation, 0–6 h)

Bias/RMSE	MYNN	SH	MYJ	YSU	BOU	QNSE	ACM2
T	−0.82/2.50	0.04/2.30	0.10/2.21	0.04/2.31	0.05/2.31	−0.37/ 2.15	0.14/2.31
DPT	−2.45/3.85	−2.60/3.84	−2.15/3.59	−2.60/3.84	−2.39/3.74	−2.05/3.53	−2.76/3.94
RH	−5.10/12.8	−7.59/13.6	−6.62/13.3	−7.60/13.6	−7.06/13.5	−5.30/ 12.6	−8.10/13.9
u comp	−0.65/2.10	−0.50/2.07	−0.70/2.15	−0.51/2.08	−0.56/2.08	−0.68/2.11	−0.55/ 2.07
v comp	−0.36/2.15	−0.32/2.11	−0.39/2.21	−0.32/2.11	−0.29/2.00	−0.35/2.17	−0.33/2.10
Wind	1.06/2.25	0.94/2.16	1.18/2.33	0.94/2.16	0.89/2.15	1.17/2.30	0.98/ 2.15

The nowcast period (daytime CBL) results are shown in Table 8 and appear notably different than those of the initial period. The best overall performing schemes are the MYJ and BouLac schemes with 4 top performing statistics. ACM2, MYNN, SH, and YSU are each superior with only a single statistic. With respect to temperature, MYJ is clearly superior with a bias of −1.65 K and an RMSE of 2.18 K. The remaining comparisons are relatively close.

Table 8 Mean bias and RMSE for 1-km ensemble members (D3) (1 km) for 1800 UTC 9 February through 0000 UTC 10 February (nowcast period, 6–12 h)

Bias/RMSE	MYNN	SH	MYJ	YSU	BOU	QNSE	ACM2
T	−2.14/2.66	−2.11/2.58	−1.65/2.18	−2.09/2.56	−1.74/2.28	−1.89/2.35	−1.87/2.39
DPT	−1.44/3.42	−1.52/3.42	−0.71/3.31	−1.50/3.41	−1.33/ 3.30	−0.89/ 3.30	−2.08/3.53
RH	0.96/6.14	0.68/5.84	1.48/6.34	0.66/5.80	0.48/5.64	1.61/6.52	−0.63/ 5.45
u comp	0.16/1.61	0.07/1.58	0.62/1.88	0.05/1.57	−0.17/ 1.55	0.98/2.15	0.06/1.59
v comp	0.42/1.78	0.46/1.75	0.39/1.84	0.44/1.73	0.40/ 1.71	0.52/1.94	0.45/1.74
Wind	−0.10/1.41	−0.21/ 1.40	0.46/1.55	−0.24/ 1.40	−0.52/1.50	0.89/1.72	−0.18/ 1.40

Model performance with respect to bias and RMSE is fairly evenly distributed for the extended forecast period (0000–1200 UTC), which represents nighttime conditions over the model domain (Table 9). Only the ACM2 scheme does not have a top performing statistic for these times. Both the MYNN and SH schemes have 3 top performing statistics. YSU and QNSE are each superior with 2 statistics. MYJ and BouLac are superior for only a single statistic. Aside from RH where MYNN is the superior scheme, the differences between the various BL/SL formulations is relatively small and typically amount to a few tenths at most.

Table 9 Mean bias and RMSE for 1-km ensemble members (D3) (1 km) for 0000 UTC 10 February through 1200 UTC 10 February (extended forecast, 12–24 h)

Bias/RMSE	MYNN	SH	MYJ	YSU	BOU	QNSE	ACM2
T	−1.39/3.23	0.30/2.88	1.01/3.06	0.29/2.87	0.41/2.88	−0.31/ 2.82	0.30/2.88
DPT	−1.33/2.77	−1.19/2.86	−1.49/2.71	−1.11/2.85	−0.94/2.71	−1.38/ 2.68	−1.30/2.82
RH	−1.26/13.7	−5.20/14.5	−8.38/15.6	−4.87/14.4	−4.78/13.9	−4.78/14.4	−5.64/14.3
u comp	0.07/1.51	0.42/ 1.34	−0.26/1.58	0.44/1.35	0.42/1.41	−0.65/1.73	0.29/1.40
v comp	0.06/1.49	0.08/ 1.41	−0.04/1.51	0.08/1.42	0.10/1.46	−0.11/1.56	0.14/1.45
Wind	0.51/1.55	0.20/1.32	0.68/1.60	0.20/ 1.32	0.34/1.38	0.93/1.74	0.32/1.34

In a similar manner, statistics were compiled for the time of the entire forecast cycle. Those results are summarized in Table 10. For the combined nowcast and extended forecast period, the SH scheme excelled in 4 measures, all of which are wind measures, followed by MYNN with 3 statistics, excelling mainly with the RH bias and RMSE measures. YSU and QNSE beat out the others in 2 statistics each and share best performance with temperature. BOU and MYJ came out on top with a single measure apiece. Despite this, the individual differences between the various schemes in terms of performance are relatively small except in isolated cases. While both the MYNN and MYJ temperature bias exceed 1 K, the others are typically a few tenths of a degree. The differences between the temperature RMSE ranges from

0.3 to 0.4 K for the other 5 schemes. While this might eliminate MYNN for consideration if we are mainly interested in surface temperature forecasts, no particular scheme would be out of contention for modeling any of the other variables, where they are the primary forecast variables of concern.

Table 10 Mean bias and RMSE for 1-km ensemble members (D3) (1 km) for 1200 UTC 9 February through 1200 UTC 10 February (complete forecast, 6–24 h)

Bias/RMSE	MYNN	SH	MYJ	YSU	BOU	QNSE	ACM2
T	−1.64/3.04	−0.50/2.78	0.12 /2.77	−0.50/2.77	−0.31/2.68	−0.84/ 2.66	−0.42/2.71
DPT	−1.37/2.99	−1.30/3.05	−1.23/2.91	−1.24/3.03	−1.06 /2.90	−1.21/ 2.89	−1.56/3.05
RH	−0.52/11.2	−3.23/11.6	−5.09/12.5	−3.03/11.6	−3.03/ 11.2	−2.65/11.8	−3.97/11.4
u comp	0.10/1.54	0.30/ 1.42	0.03 /1.68	0.31/ 1.42	0.22/1.46	−0.11/1.87	0.22/1.46
v comp	0.18/1.58	0.20/1.53	0.10 /1.62	0.20/ 1.52	0.20/1.54	0.10 /1.69	0.25/1.55
Wind	0.31/1.51	0.07/ 1.35	0.61/1.59	0.06/ 1.35	0.05 /1.42	0.92/1.73	0.15/1.38

Depending on the time period of interest, the time of day, and the primary meteorological variable(s) of primary concern to the end user, an individual scheme can be chosen to minimize the error(s) of primary interest. If we are primarily interested in surface wind forecasts, the SH scheme would be preferable. However, in most cases the differences between each scheme are typically very small. Except in clear cases of under-performance in a particular scheme or schemes (for example, the MYNN scheme for temperature during an entire 24-h cycle), any of the other schemes remain candidates for inclusion in a preconfigured modeling system deployed in the battlefield without a priori knowledge of the location of deployment, and the time of the simulation.

The statistics shown in our analyses do not define a single superior scheme; rather, the analyses show that the best overall scheme is a function of time of day and of the meteorological variable of interest. The standard measures of forecast quality (RMSE and bias) commonly vary insignificantly between the schemes tested. To choose the best overall scheme, we assigned rankings (from 1 through 7) for each of the 3 independent variables (T, DPT, and wind) and averaged the rankings for the nowcast, extended forecast, and complete forecast periods. The data assimilation period is not included in this analysis. Furthermore, since the bias calculation potentially includes large swings in the error that may be offset by one in the other direction, we chose to evaluate the overall relative performance of the schemes using only the RMSE.

The results of this determination are shown in Table 11. The mean RMSE and the standard deviation of the RMSE for the 7 schemes tested are shown. In addition,

the rankings for T, DPT, and wind speed are shown. To obtain a representative ranking that does not over account for wind, we eliminated the u and v components and RH from our analysis. Each scheme is ranked 1–7 for the nowcast, extended forecast, and complete forecast periods.

Table 11 Summary RMSE statistics (mean and standard deviation) and relative ranking of each of the PBL scheme configurations we tested for the nowcast, extended forecast, and total forecast periods. Overall ranking for each scheme (COMP) is a composite of T, DPT, and wind speed rankings. Temperature and DPT units are in Kelvin and wind speed in ms^{-1} .

Root Mean Squared Error (RMSE) Statistics and Ranking										
Hours	Field	MEAN	STDEV	MYNN	SH	MYJ	YSU	BouLac	QNSE	ACM2
06 thru 12 (nowcast)	TEMP	2.43	0.18	7	6	1	5	2	3	4
	DPT	3.38	0.09	5	5	3	4	1	1	7
	WIND	1.49	0.12	4	1	6	1	5	7	1
	COMP			7	5	3	2	1	4	5
12 thru 24 (extended)	TEMP	2.95	0.15	7	3	6	2	3	1	3
	DPT	2.77	0.07	4	7	2	6	2	1	5
	WIND	1.47	0.16	5	1	6	1	4	7	3
	COMP			7	5	6	1	1	1	4
06 thru 24 (full fcst)	TEMP	2.77	0.13	7	6	4	4	2	1	3
	DPT	2.97	0.07	4	6	3	5	2	1	6
	WIND	1.47	0.14	5	1	6	1	4	7	3
	COMP			7	5	6	3	1	2	4

For the nowcast period, corresponding to daytime, the BouLac scheme performs best. For the extended forecast, corresponding to nighttime, YSU, BouLac, and QNSE tied for best performance. Overall, for the entire forecast period, BouLac performs best for our data. The ensemble mean RMSE for temperature varies between 2.4 and 3.0 K for our data. The DPT RMSE is slightly larger and ranges from 2.8 to 3.4 K. The mean RMSE for wind velocity is nearly constant at approximately 1.5 ms^{-1} . Note that the standard deviation of the RMSE is quite small, indicative of the lack of spread between the various PBL schemes we tested.

All of the results presented thus far have focused on the model statistics derived from the innermost 1-km grid domain since it is the domain of primary interest. Since the 1-km domain resides entirely within the 3-km middle domain, we were interested in what additional accuracy we might obtain by using such a fine-scale grid over the coarser 3-km grid. Table 12 shows a comparison of the RMSE and bias over the 1- and 3-km grids for a single meteorological variable, temperature. The data included for the 3-km domain are restricted to those points that are members of the 1-km domain only. This table shows values for a single time in the nowcast period of the simulation; namely, 0000 UTC (1600 PST) on 10 February (the 6-h forecast). All bias and RMSE estimates were close for both the grid spacing

and for the individual BL/SL schemes within a couple tenths of a degree Kelvin. At nighttime, before sunrise, the model underestimates the surface temperature by about 2 K. Curiously, when we look at grid spacing, the coarser 3-km resolution grid produces very slightly better results. However, the improvement is small both with regard to the model bias and RMSE. This may or may not be statistically significant and may be related to geolocation errors, the limitations of a grid-to-point verification method, or other interpolation assumptions used when interpolating the model data to the observation locations.

Table 12 Temperature bias (2m AGL) and RMSE for 1- and 3-km ensemble members (D3/D2) (1 km/3 km) validation time: 0000 UTC on 10 February (12-h simulation time, and 6-h lead time)

Error stat	MYNN	SH	MYJ	YSU	BOU	QNSE	ACM2
Bias	-2.31/-2.21	-2.32/-2.26	-2.34/-2.34	-2.29/-2.25	-1.99/-1.97	-2.61/-2.62	-2.11/-2.09
RMSE	2.78/2.57	2.74/2.60	2.68/2.66	2.72/2.60	2.46/2.34	2.91/2.90	2.57/2.46

While the 3-km grid very slightly appears to be better, it may not necessarily indicative of an increase of performance with the coarser grid. Further studies may be deemed necessary, but that is beyond the scope of this report.

5.2 Talagrand Diagrams

To provide an estimate of the utility of the 7 member PBL/SL scheme combinations for producing a valid ensemble that could provide useful model statistics and quantification, Talagrand diagrams (Hamill 2001; Wilks 2011), also known as rank histograms, were produced for this study. The shape and distributions shown by these diagrams represent a measure of the validity of an ensemble of forecasts and tell us about the relationship of the forecasts to the observed data. This information includes model bias and error and the spread of the set of forecasts. Essentially, the “ranks” represent the position within the ensemble where the observed value is positioned. So a rank of 3 would indicate that the observed value is in position 3 of the ensemble members (sandwiched in between ensemble members 2 and 3), and so forth. A rank of 8 would indicate that the observed value occurs after all 7 members of this ensemble. So in this case all 7 members produce values that are lower than the observed value. It is clear, then, that all of the ensemble members are under-forecasting the value. Rank histograms were produced from the T, DPT, RH, wind speed, and u and v component fields output from WRF and supplied to the MET package for this analysis.

The results of this analysis for 00Z 10 February are shown in Figs. 22–27. The 2m temperature histogram is shown in Fig. 22. Very few data are represented in bins 1–7, and a majority of the measurements occur in bin 8. This type of plot is produced when the observed value is higher than nearly all of the ensemble forecasts. Thus, in our case, the model is underestimating the surface temperature at 00Z, which is late afternoon in southern California. Since very little data appear in the other bins, the WRF model consistently underestimates the late afternoon surface temperatures for the 7 PBL/SL schemes considered in this analysis. The observed temperature is consistently higher than the ensemble members and independent of which scheme we choose for the PBL/SL physics.

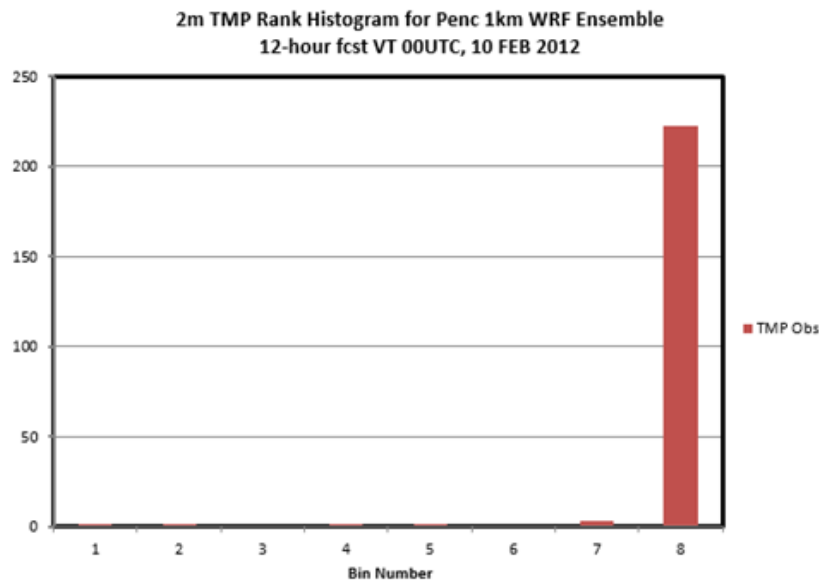


Fig. 22 WRF D3 (1-km) ensemble rank histogram produced for 00Z 10 February 2012, surface (2m) temperature

The Talagrand diagram for DPT (Fig. 23) is notably different. In this case we obtain a classic U-shaped plot, which is produced when there are a sufficient number of low and high biases but the individual members do not spread out enough to provide an adequate statistical measure of spread within the ensemble. Although there is some spread in the ensemble members, the spread is small, resulting in relatively few samples in the intermediate bins and a majority lying at the extreme bins. While this represents an improvement in the quality of the spread produced by the ensemble, it remains inadequate, as the idealized plot would be relatively flat through all 8 bins (7 members plus the observations).

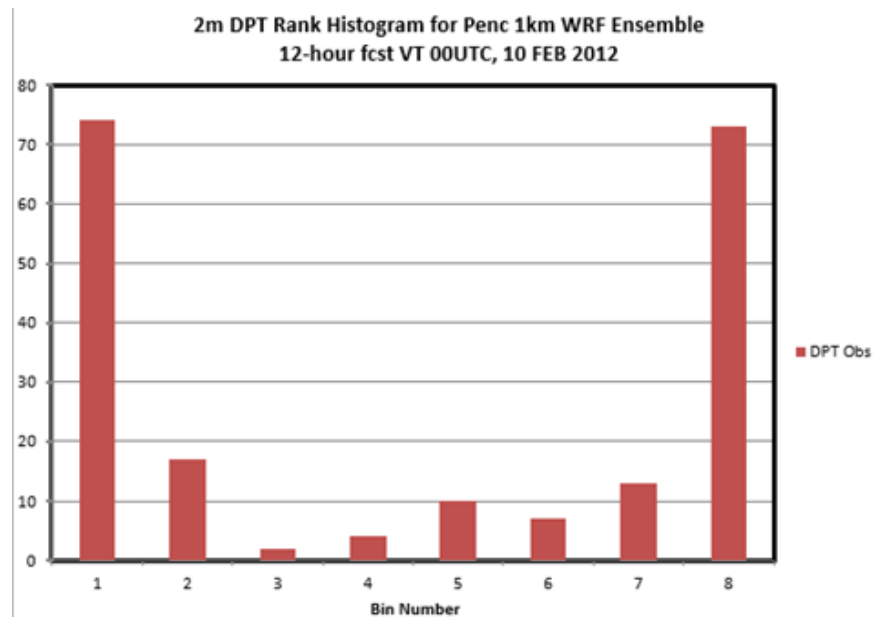


Fig. 23 WRF D3 (1-km) ensemble rank histogram produced for 00Z 10 February 2012, surface (2m) DPT

Similar results were obtained when we looked at the RH (Fig. 24), u component (Fig. 25), v component (Fig. 26), and wind velocity (Fig. 26). The RH Talagrand diagram also indicates that the model is biased high, at least at 00Z. This is physically consistent with the model bias toward underestimating the surface temperature, as we have seen earlier. The wind components and the velocity rank histograms also show that the model errors are widely spread.

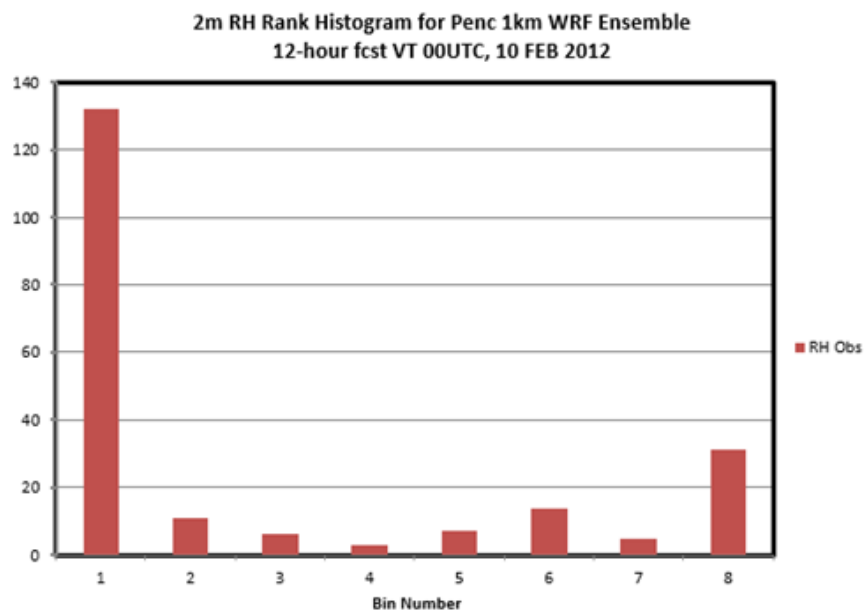


Fig. 24 WRF D3 (1-km) ensemble rank histogram produced for 00Z 10 February 2012, 2m (surface) RH

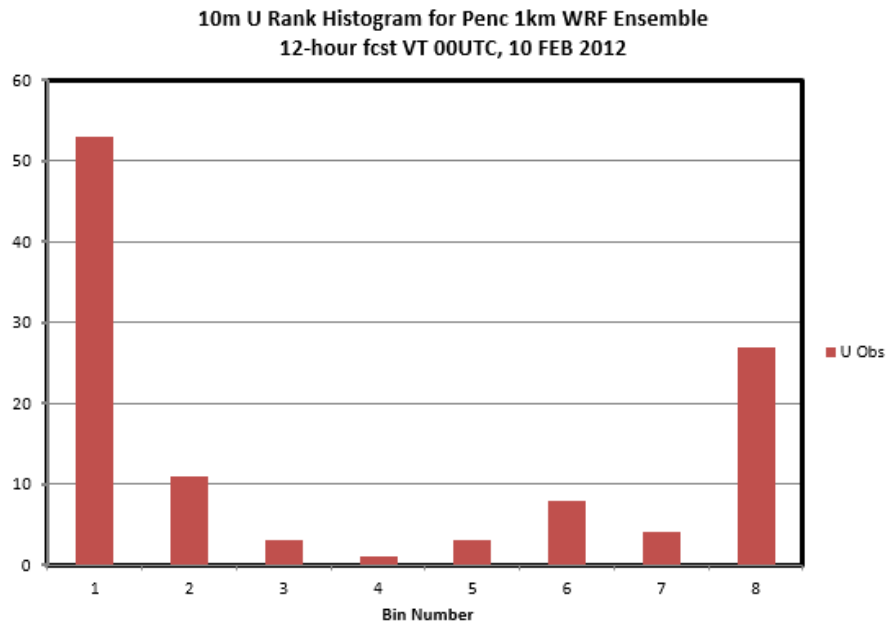


Fig. 25 WRF D3 (1-km) ensemble rank histogram produced for 00Z 10 February 2012, 10m u component of wind

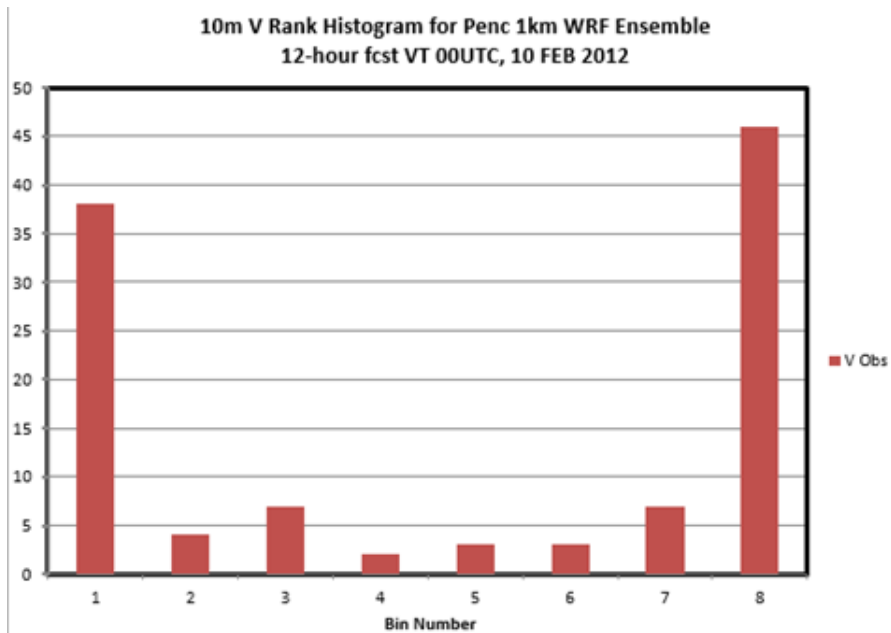


Fig. 26 WRF D3 (1-km) ensemble rank histogram produced for 00Z 10 February 2012, 10m v component of wind

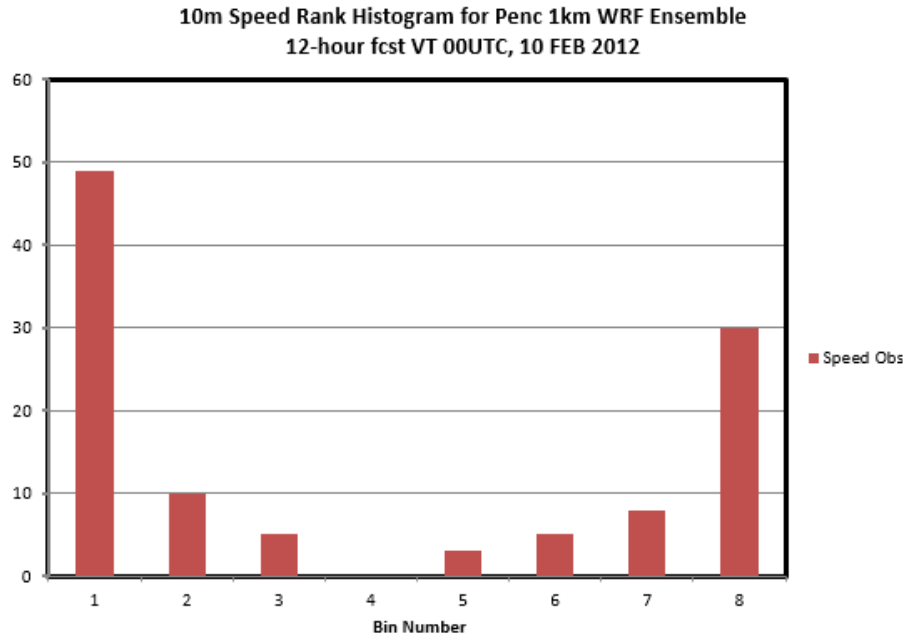


Fig. 27 WRF D3 (1-km) ensemble rank histogram produced for 00Z 10 February 2012, 10m wind speed

The characteristic U-shaped patterns are consistent with an ensemble whose members do not have sufficient spread, contain a mixture of low and high biases, and it indicates that the use of PBL/SL schemes alone does not constitute a “good ensemble”. A “good ensemble” has sufficient spread among its members such that the distribution is within the envelope of expected values for a given probability distribution defined by the validation observation(s). While we have shown only the results from the 00Z analysis, analyses from other times in this run set are similar. Therefore, we conclude that using an ensemble based solely on the choice of PBL/SL scheme combination, is, in itself, insufficient in creating a valid ensemble for the purpose of nowcasting with this version of WRF, and the current dataset (i.e., a single case day), geography, and synoptic conditions.

5.3 Horizontal Variation in Surface (2m) Temperature

The interior domain used in this modeling study contains a large variation in surface terrain, vegetation, land use, and elevation. Furthermore, there is a wide variation in climatic regimes since part of the domain is under the influence of maritime flow from over the ocean as well as arid or semi-arid inland locations. While the western extent of the domain is located at or near sea level, there are significant elevated regions to the east with mountainous terrain.

The terrain and elevation of the 1-km domain are shown in Fig. 28, along with the locations of most of the persistent weather observations used to validate the model

data for this study. Note that the only observations used in this study are those that persist throughout the entire time of integration. The elevation varies from below sea level to over 3 km, reflecting the complex terrain and the associated complex flow fields associated with this type of elevation variation over a short distance. The elevation varies nonlinearly and roughly west (low) to east (high). In addition, there is a high degree of small-scale structure, which necessitates the use of finer-resolution grids to resolve local flows. With a 1-km grid in the inner nest, this translates to the ability to resolve approximately 5-km-scale phenomena. With some of the topographic features smaller than 5 km, the model is inherently incapable of resolving them all, and this introduces some additional degree of model uncertainty, as we would be unable to represent the smallest scales of motion using this model resolution. Because of the limited computational resources in a field-deployed modeling system, we needed to effect a compromise between the need to resolve finer-scale meteorological phenomena and the limited resources available in the forward-deployed environment. While we were not as restricted in this study, we wanted to mimic the resources present in the typical battlefield deployment the best we could.

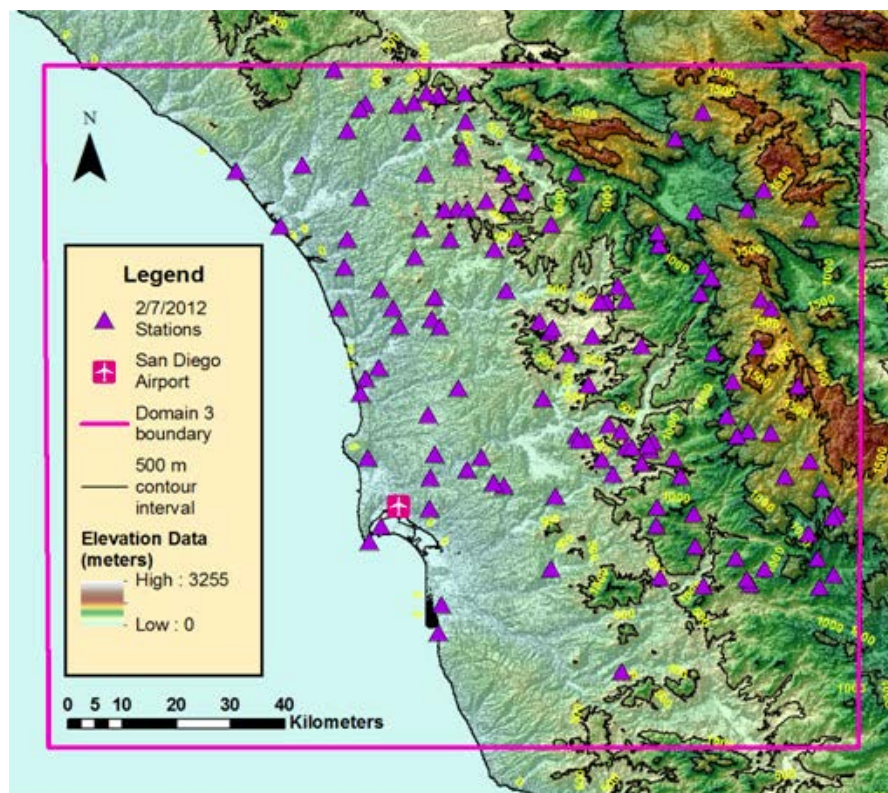


Fig. 28 Terrain and location of the weather stations that contain persistent data for a related study over the innermost domain (1-km grid) created from a US Geological Survey digital elevation model with a resolution of 1/3 arc-second (Smith et al. 2015). A subset of these observations is used based on persistence through model integration.

The mean 2m surface temperature and terrain height across the inner domain at 0000 UTC 10 February is shown in Fig. 29. The mean value is taken by averaging paired observations over the 7 members of the ensemble. As expected, the highest temperatures are found at the lower elevations with the exception of the region immediately surrounding the coastline. The interior lowlands have the highest temperature. Conversely, the coolest temperatures are found at the higher elevations. Over the domain, the elevations range from sea level to well over 1500 m. Thus, there is a direct correlation between the elevation and surface temperature. Over such a limited domain, there is little relationship between temperature and latitude. A secondary area of temperature maximum lies to the lee of the mountains at the eastern extremity of the model domain.

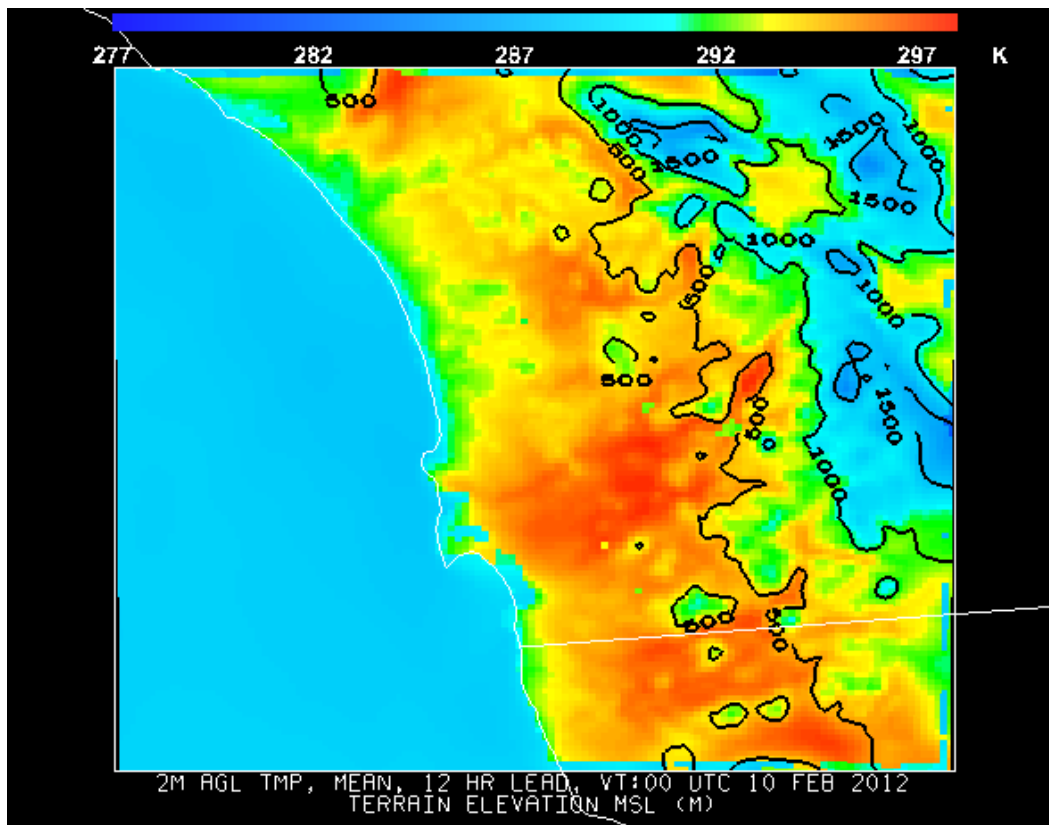


Fig. 29 Ensemble mean surface (2m AGL) temperature (Kelvin) over the innermost model domain (D3, 1-km grid) at 0000 UTC 10 February 2012

The cooler temperatures near the coastline are a reflection of the relatively cool waters of the eastern Pacific and the marine BL. These temperatures are not as cool as those associated with the higher terrain. The range in surface (2m) temperature extends from approximately 281 to 297 K.

The ensemble maximum at 0000 UTC 10 February is plotted in Fig. 30. Here the range extends from approximately 284 to 298 K. The difference between the mean and maximum values within the members is approximately 2 K. This does not represent a large spread in the values compared with the ensemble mean. The same relationships exist that were found for the plot of the means over the model domain. There are no notable deviations from the previous analysis of the ensemble mean values. However, the lowest ensemble maximum in the 0000 UTC temperature forecast occurs leeward of the mountains toward the eastern edge of the model domain, so the mean surface temperature and maximum surface temperature are more comparable across this part of the domain.

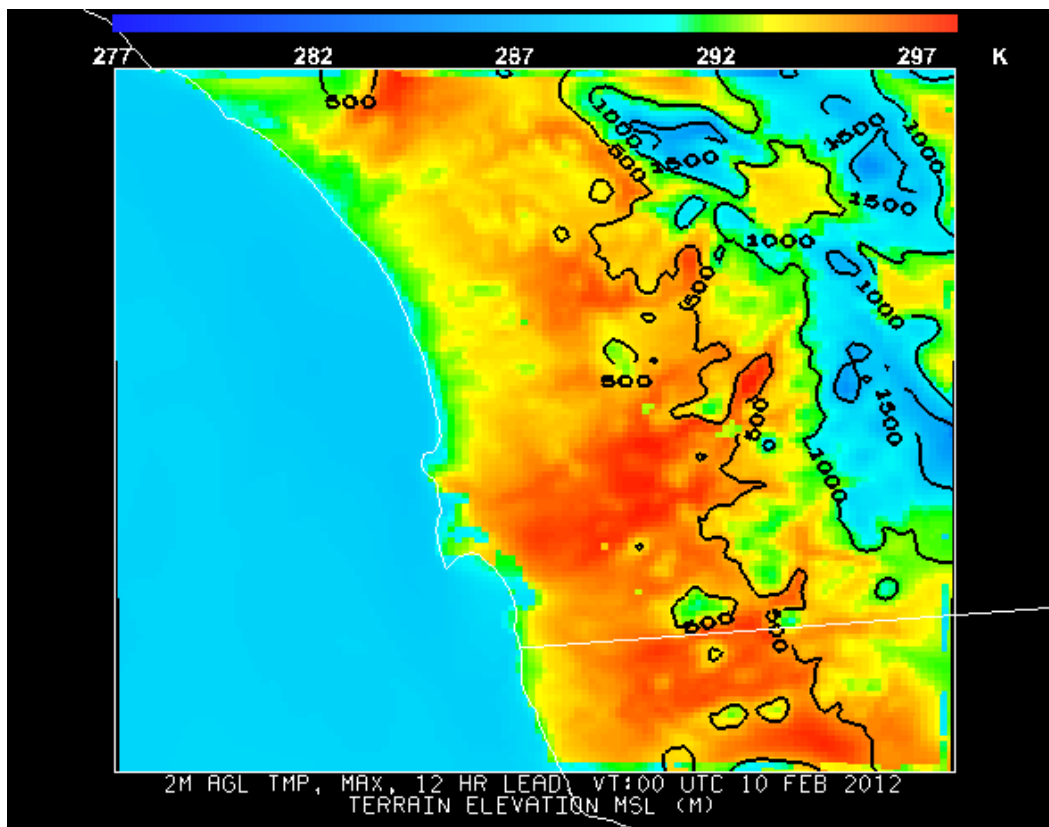


Fig. 30 Ensemble maximum surface (2m AGL) temperature (Kelvin) over the innermost model domain (D3, 1-km grid) at 0000 UTC 10 February 2012

The ensemble minimum surface temperature at 0000 UTC 10 February is plotted in Fig. 31. The range extends from approximately 277 to 297 K. The primary difference between this field and the mean is that the warmer temperatures extend closer to the coastline. The other features are largely consistent with the previous plots. The largest horizontal variation in the ensemble minima occurs near Point Loma, which is a peninsula surrounding San Diego Bay with an area of approximately 18 km². This is an area where it is difficult for the model to resolve

small-scale flow associated with the geographic feature since convergent and/or divergent flows can be triggered by the geography. As in the earlier plots, the boundary issues become apparent over land to the northern, eastern, and southern edges of the domain.

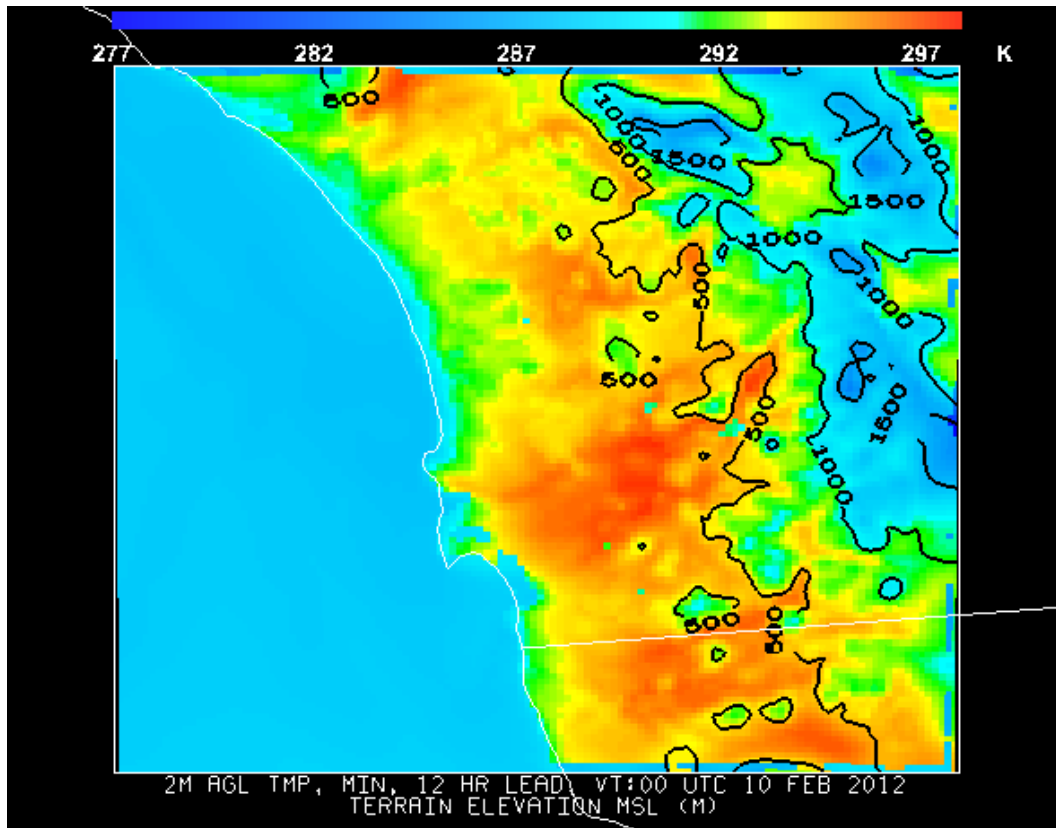


Fig. 31 Ensemble minimum surface (2m AGL) temperature (Kelvin) over the innermost model domain (D3, 1-km grid) at 0000 UTC 10 February 2012

Figure 32 shows the ensemble range over the innermost (1-km) grid at 0000 UTC 10 February. Most of the variability between ensemble members occurs in the range from 0.1 K to a few tenths of a degree Kelvin. Very little variation occurs between ensemble members both at inland locations and over the Pacific Ocean, which is consistent with previous analyses (Smith et al. 2015, 2016a,b) showing very limited spread among the ensemble members for temperature. There is a hint of a slightly larger variance between members in locations near the coastline. This is to be expected since any variation in the land–sea breeze circulation may explain these differences. The range in surface temperature between ensemble members is also larger in the San Diego area, as evidenced by the increased light blues and the green, yellow, and red areas located in and to the north of the city. These colors correspond to 1.6 to 2.2 K. There is no clear relationship between the ensemble range and elevation over the inner (D3, 1-km) model domain.

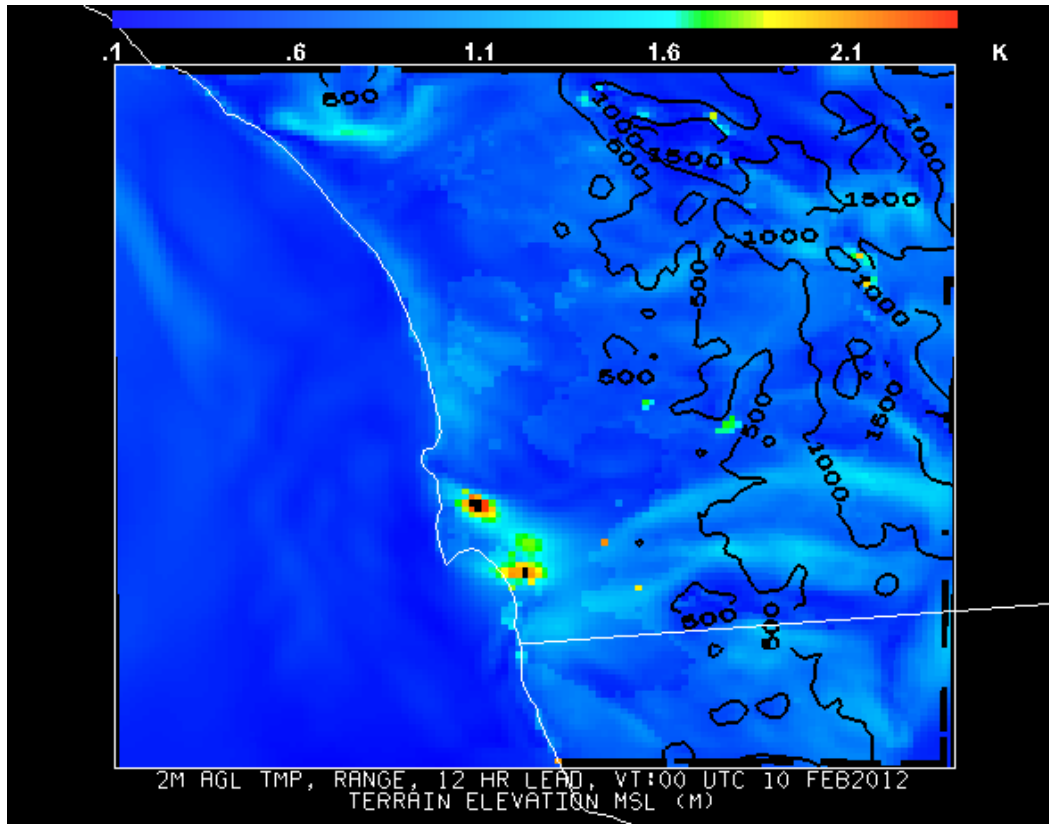


Fig. 32 Ensemble range, surface (2m AGL) temperature (Kelvin) over the innermost model domain (D3, 1km grid) at 0000 UTC 10 February 2012

The ensemble standard deviation is shown in Fig. 33. The standard deviation of temperature is small, less than 0.2 K over nearly the entire domain, which is consistent with the very small spread among the 7 model members evident from prior analyses. For the most part, the standard deviation of the surface temperature runs from 0.1 to 0.2 K. The smallest standard deviations over land occur northeast of Point Loma and the San Diego Bay, which lies immediately to the south of the peninsula, directly north of San Diego, toward the northern edge of the domain and over higher terrain in the northeast quadrant of the model domain. There is no distinct pattern in the standard deviation, as the smallest spread in the members occurs largely over water as would be expected in a marine environment. The higher elevations to the north also show a reduced spread in the model-predicted temperature. However, there is no clear relationship between standard deviation and terrain elevation, in general, over the model domain. The largest standard deviations are observed over land, as one would expect, with the largest standard deviation being only about 0.2 K, which defines the uppermost limit in the figure legend. The coastline is readily observable as a sharp difference in the standard deviation. There is little spread in the ensemble members.

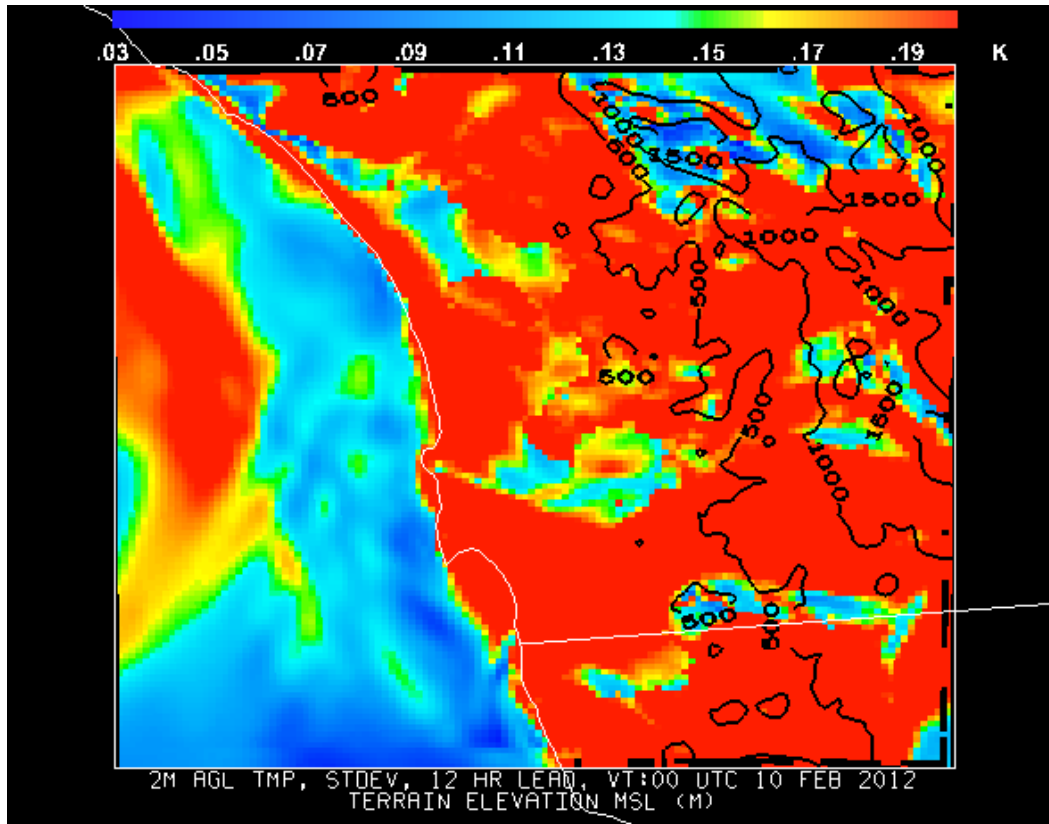


Fig. 33 Ensemble standard deviation of surface (2m AGL) temperature (Kelvin) over the innermost model domain (D3, 1-km grid) at 0000 UTC 10 February 2012

The ensemble mean plus and minus 1 standard deviation (σ) values for the innermost domain for 0000 UTC 10 February are shown in Figs. 34 and 35, respectively. The ensemble surface temperature mean plus 1 σ surface temperature ranges from approximately 284 to 298 K. The ensemble mean surface temperature minus 1 σ varies from 280 to 298 K over the model domain. These plots are useful in defining the central lobe of the spatial distribution of temperature entire model domain, encompassing roughly 68% of the variation, the range over which most of the ensemble variance extends over the innermost (D3) model domain, and explaining the horizontal distribution of this measure. The most likely solution normally lies in between these 2 estimates of temperature. There are no remarkable features in either plot, and the distribution closely mimics that of the ensemble mean surface temperature, the ensemble mean maximum, and ensemble mean minimum plots shown earlier (Figs. 29–32).

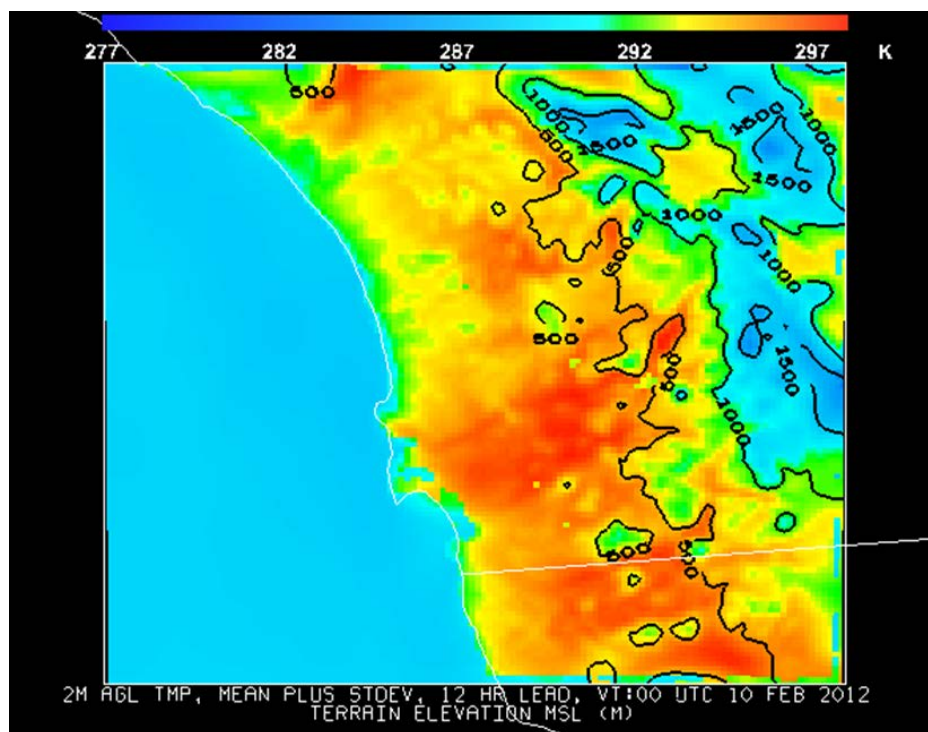


Fig. 34 Ensemble plus 1 σ of surface (2m) temperature (Kelvin) over the innermost model domain (D3, 1-km grid) at 0000 UTC 10 February 2012

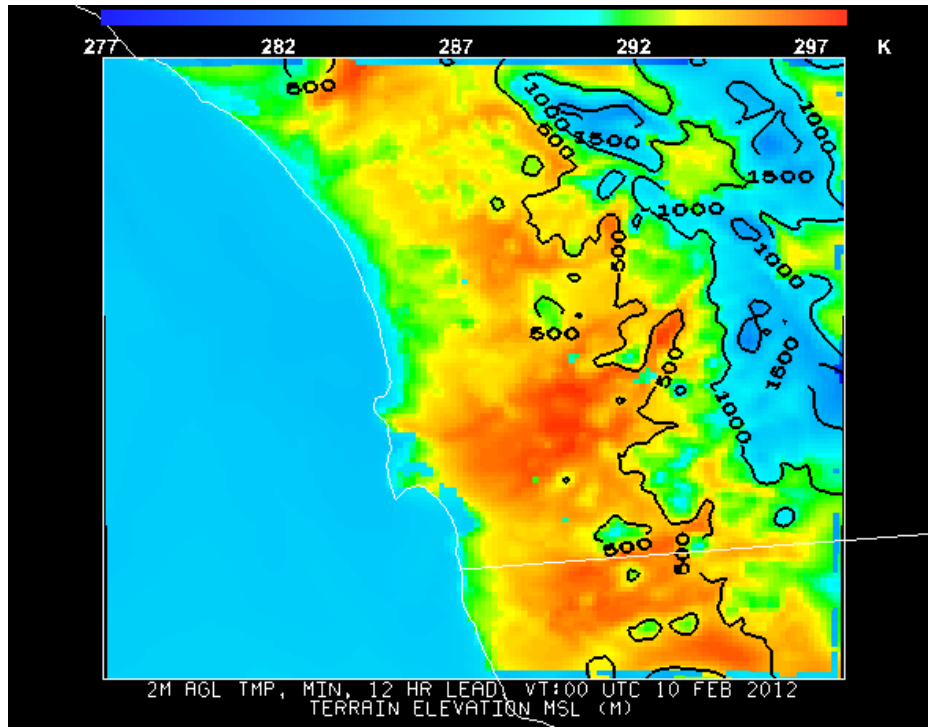


Fig. 35 Ensemble minus 1 standard deviation of surface (2m AGL) temperature (Kelvin) over the innermost model domain (D3, 1-km grid) at 0000 UTC 10 February 2012

Basically the mean plus and minus 1 σ plots should describe about 68% of your data if the data are Gaussian. Essentially, this information is useful since it would give you the main lobe of the data. The utility of these plots might be constrained for the domain as a whole because of potential multimodal distributions and other model variables like WIND, UGRD, and VGRD. Nevertheless, the plots show the extent of the main lobe of where the ensemble spread is located.

Smith et al. (2015) looked more closely at the distribution of the temperature error and its contribution to the model performance versus a number of explanatory terrain variables, and conclude that elevation is the most important. It is intuitive that the temperature distribution over the domain will be strongly related to the elevation, and our results support this. Smith et al. could not formulate a conclusion as to the diurnal pattern in how elevation relates to temperature. Even over this small inner domain, there is large variation in the elevation and of land surface type.

While the Smith et al. (2015) study provides a way ahead to investigate subdomain processes, atmospheric response, and model performance, that research has not extended past the initial study. Furthermore, we have not performed a complete analysis on the subdomain temperature distribution because we lack the appropriate GIS tools. The standard postprocessing software included in the WRF Modeling System and the MET package we used in this study generally processes model statistics only over an entire domain. While it is possible to perform a detailed analysis at the subdomain level, we are constrained to using polygons, for example, a nested grid, to downselect our data, which makes selection of subregions far less precise and limited when using specific selection criteria. The use of GIS tools makes selection based on land use, terrain, elevation or any other selection criteria much easier.

5.4 PBL Depth

Another useful meteorological quantity aside from surface values of temperature, moisture, and wind is the PBL depth, which is useful for some meteorological sensors and weapons systems since this measure gives an indication as to where enhanced turbulence due to convective mixing is expected, and how deep mixing would be expected to dilute materials released into the atmosphere. Optical systems are also highly affected by atmospheric turbulence, including mixing and the formation of BL clouds. Turbulence also affects radio propagation and the performance and safety of unmanned aerial vehicles, so knowledge of the PBL depth defines the layer where degraded performance may be expected. Our dataset, unfortunately, has very few data on which to validate our model PBL depth forecasts. Only one routine upper air observation station, at Miramar Marine Corps Air Station (KNKX), is located within the inner (D3) domain (Fig. 36).

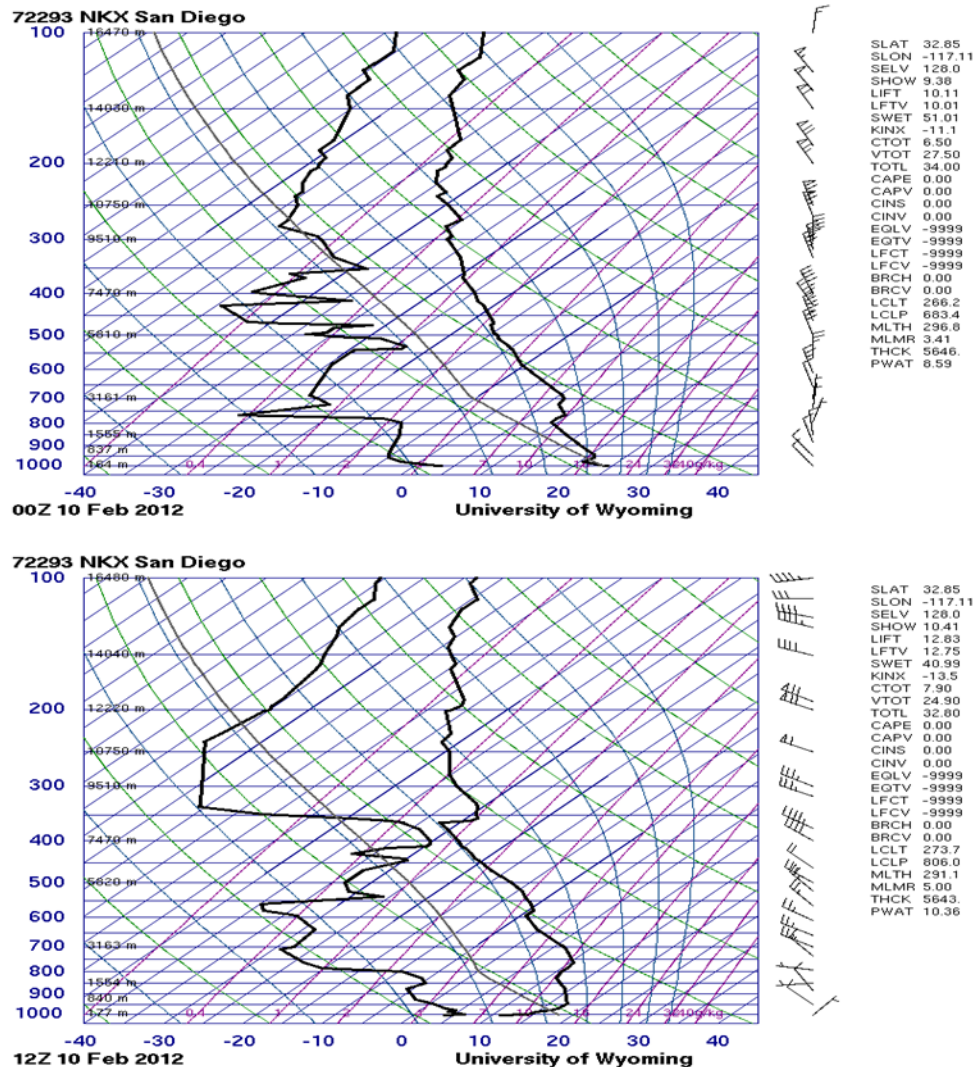


Fig. 36 Atmospheric soundings at KNKX, San Diego (top) 0000 UTC 10 February 2012 and (bottom) 1200 UTC 10 February 2012 (Courtesy of the University of Wyoming: <http://weather.uwyo.edu/upperair/sounding.html>.)

The sounding for 1200 UTC 9 February (0400 PST) at the beginning of the simulations is not shown, as it was incomplete. However, the temperature profile shows a strong ground-based inversion that extends to approximately 980 mb. A slightly less than adiabatic temperature profile then extends to approximately 750 hPa with light easterly winds through the layer. The next sounding at 0000 UTC 10 February (1600 UTC), shown in the top image of Fig. 36, represents late afternoon conditions, showing a thin superadiabatic layer from the surface to about 980 mb. From the surface to the base of the capping inversion at 800 mb, winds are light, between 2.5 and 5.0 kt. The last sounding at 1200 UTC 10 February (0400 PST) shows the strong surface inversion associated with the nocturnal BL, extending to 950 mb, with light (~2.5 kt) winds up to approximately 750 mb.

The diurnal variation in modeled PBL depth at a point near the center of the inner grid (32.9° latitude, 117.1° longitude) is shown in Fig. 37. The time runs from left to right and begins at 1200 UTC 9 February and extends to 1200 UTC 10 February. Aside from the initial bump around 1300 UTC, the nocturnal PBL is present until growth of a daytime CBL begins around 1500 UTC. The BL growth depicted by the model is nearly linear and reaches a maximum depth at around 2100 UTC (1400 PST local time). The CBL then generally erodes, again, almost linearly until about 0600 UTC 10 February. After that point a very shallow nocturnal PBL is evident until the end of the simulation.

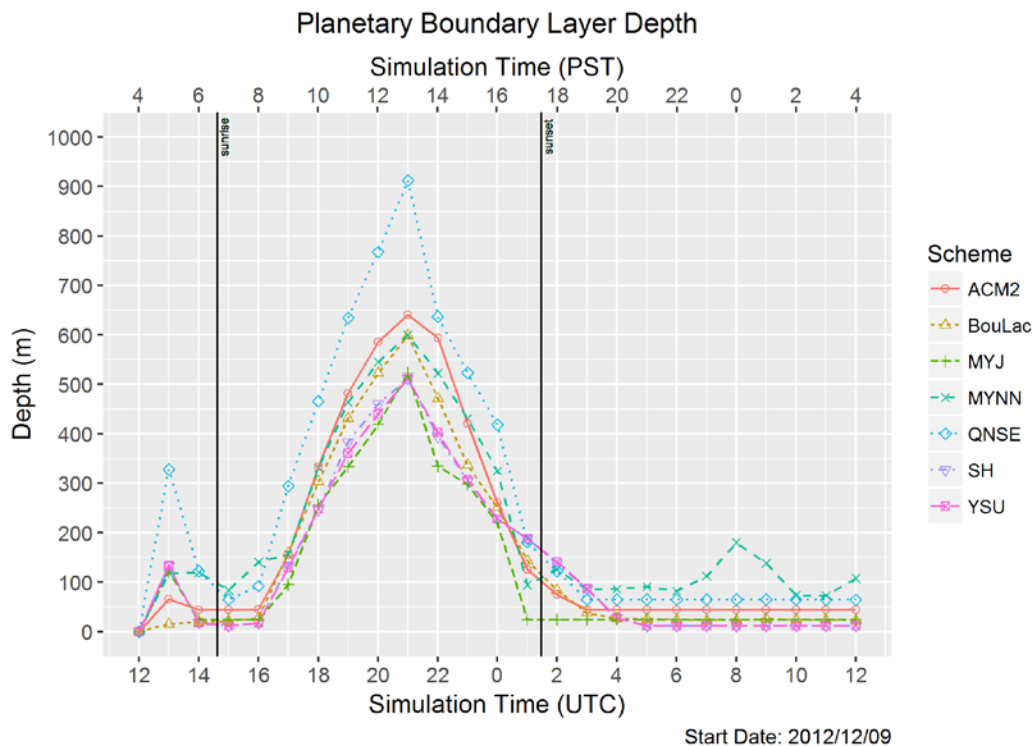


Fig. 37 Temporal evolution of PBL depth for the 7 PBL schemes tested over the center point of innermost (D3, 1-km) domain

Qualitatively, this behavior is consistent with what is expected in the evolution of the PBL over land. However, it is not possible to follow the growth and decay with the available upper air observational data, largely because we have only one available site to use in the analysis and only point measurements (soundings) at regularly scheduled times (i.e., 0000 UTC and 1200 UTC at KNKX). However, the general behavior is captured by the model.

When comparing PBL depths forecast by different BL schemes it is important to be aware that the methods used to diagnose the PBL depth vary among the schemes. Thus, it is possible for 2 schemes to forecast the same temperature and moisture profiles and yet diagnose a different PBL depth. Differences in PBL depth among

experiments are then a combination of differences in the forecast structure of the atmosphere and differences in the methods used to diagnose the PBL depth. Two general methods are used to estimate the depth of the PBL: one based on the Richardson number and the other based on the vertical profile of temperature. The thresholds for detection vary among the individual schemes. Reen et al. (2014b) explored this issue in more detail.

We did perform a limited qualitative analysis of the variation in the PBL depth by the model over the innermost (D3) domain (Figs. 38–43). Each figure shows the D3 PBL depth at 0000 UTC (1600 PST) 10 February 2012 for one of the 7 experiments and includes contours of the terrain height (in meters).

Figure 38 shows the horizontal variation of the PBL over the innermost model domain at 0000 UTC 10 February for the MYNN PBL scheme. As expected, the lowest PBL depths are found over the cool waters since the daytime surface temperature increase is greater over the land than over the water. Where surface temperature is the highest, we find the deepest PBLs. Areas removed from the coast show greater PBL depths. The enhanced surface mixing caused by disturbed air flow over the higher terrain also results in greater PBL depths. What we can tell from these data is that the model appears to be reasonably reproducing the basic physics that controls the depth of the PBL. That is, the surface heating and mixing combine to contribute to greater PBL depths with the expected geographical distribution. The maximum depths presented by the model data are in the vicinity of 1.0 to 1.2 km, which is more or less typical of the surface heating expected during the wintertime over Southern California. Near the coastlines, and over water, the PBL depths are from a few tens of meters to a couple hundred meters. These values are again consistent with the formation of a stable marine BL, which is expected over the cold waters of the eastern Pacific (Angevine 2006). The small-scale structure evident over the eastern half of the domain reflects the effects of the higher terrain in generating mixing from the surface to drive the PBL and the enhanced entrainment of air from above the capping inversion in these situations.

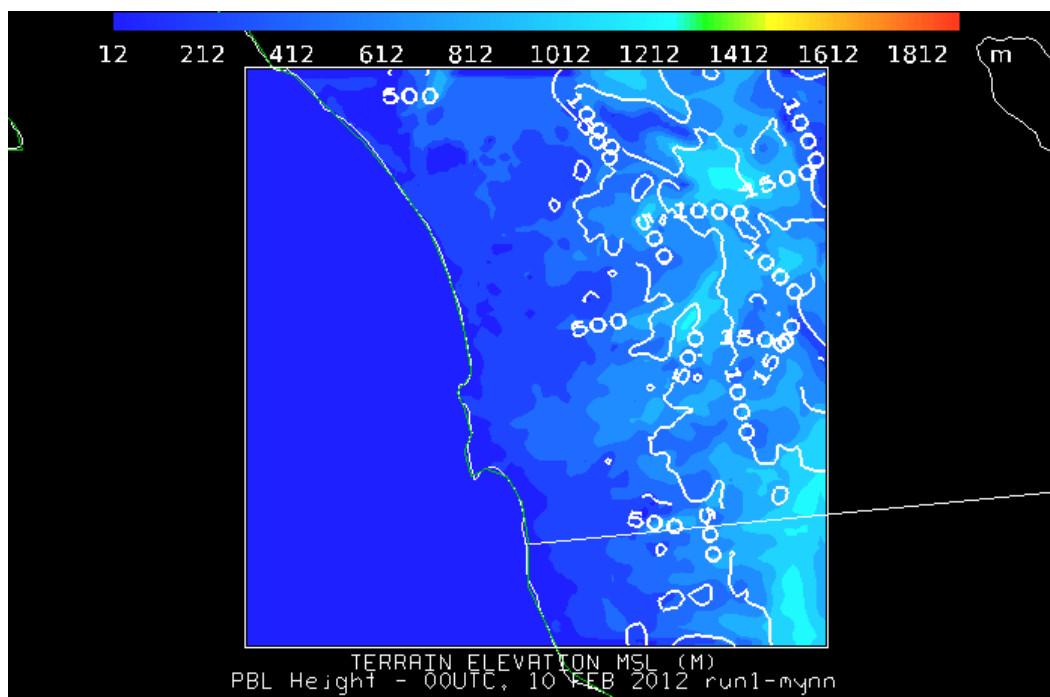


Fig. 38 PBL depth (meters) valid at 0000 UTC 10 February 2012 for the MYNN PBL option

The results from the SH scheme are shown in Fig. 39. The primary difference between the MYNN and SH schemes is that SH appears to produce slightly shallower daytime CBL depths. There also appears to be less small-scale variation using this scheme than was found with MYNN, and some of the terrain effects appear to be smoothed out over the eastern portion of the model domain. Slightly more fine-scale structure appears when using the MYJ PBL scheme (Fig. 40). Particularly toward the southeastern portion of the model domain, greater PBL depths are found. However, over the majority of the model domain, the PBL depths are consistent with those of SH. The YSU scheme PBL depth (Fig. 41) looks very similar to that of SH, which is expected since the SH scheme is built upon the YSU scheme with some small differences. Comparison of Fig. 39 with Fig. 41 shows some of the similarities. There is very little difference in the horizontal variation of the PBL depth in these scenarios. The MYJ scheme (Fig. 40) shows finer-detailed structure in the southeastern corner of the domain than the other schemes. The BouLac-derived PBL depth (Fig. 42) appears similar to the MYNN case. The horizontal distribution over the model domain and the minimum and maximum values are close to identical for the BouLac and MYNN schemes.

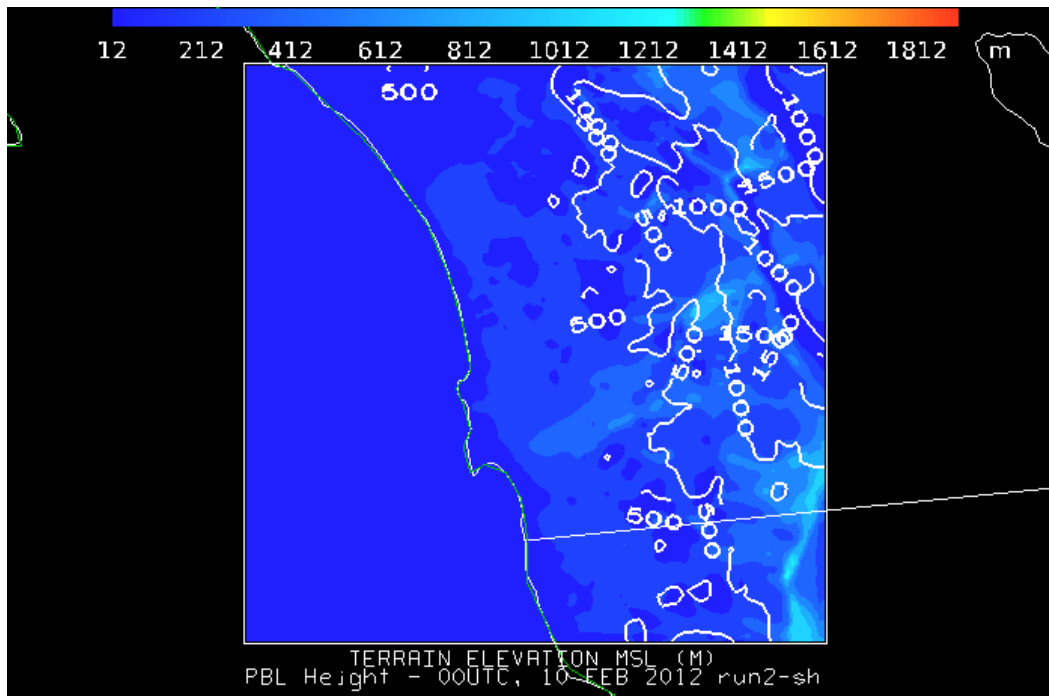


Fig. 39 PBL depth (meters) valid at 0000 UTC 10 February 2012 for the SH PBL option

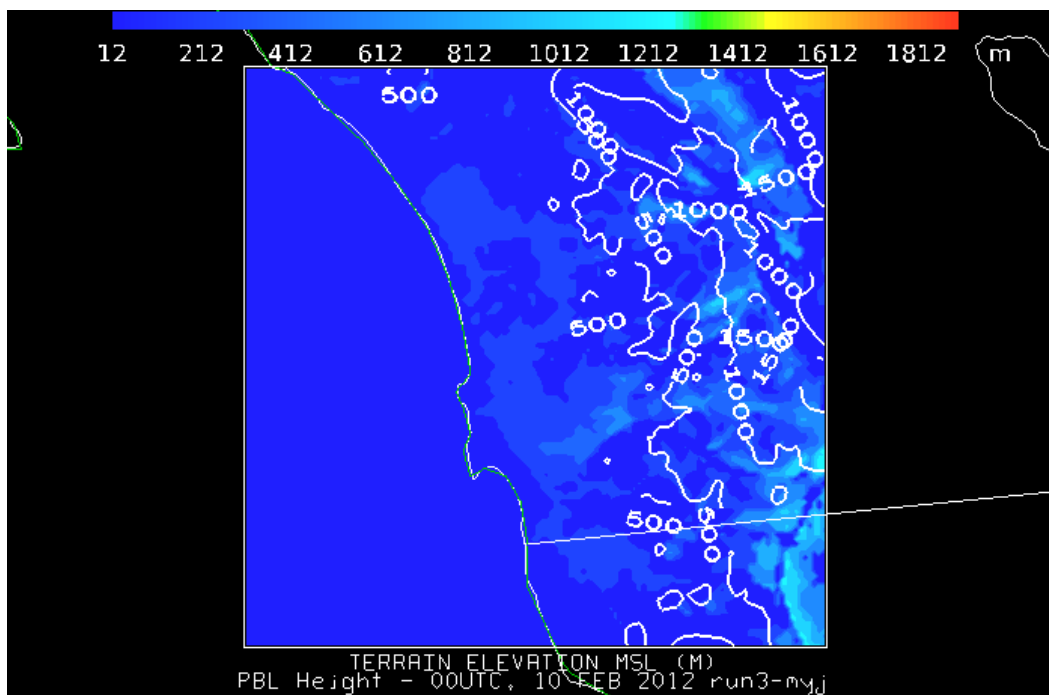


Fig. 40 PBL depth (meters) valid at 0000 UTC 10 February 2012 for the MYJ PBL option

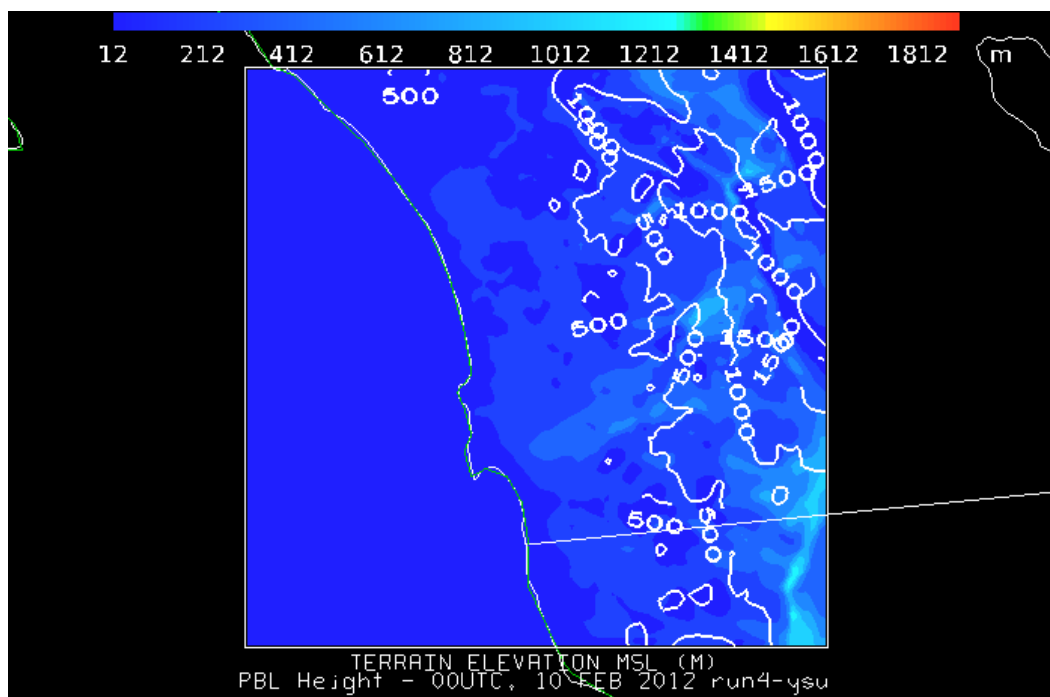


Fig. 41 PBL depth (meters) valid at 0000 UTC 10 February 2012 for the YSU PBL option

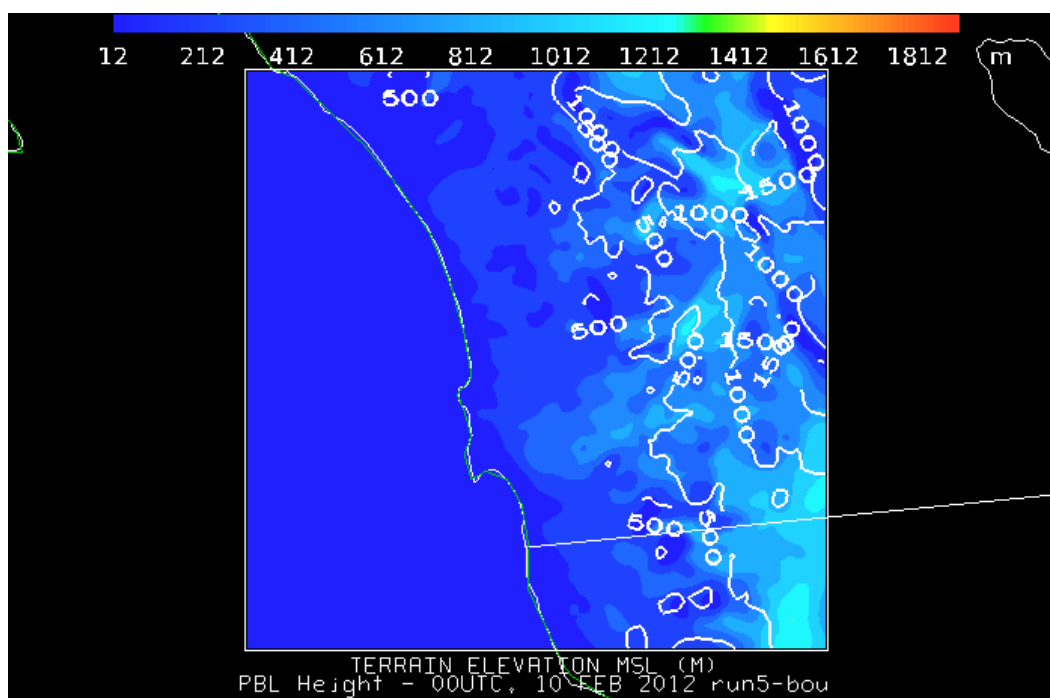


Fig. 42 PBL depth (meters) valid at 0000 UTC 10 February 2012 for the BouLac PBL option

In contrast with the MYJ, MYNN, SH, and YSU schemes, the QNSE scheme produces significantly greater PBL depths, particularly over the southeastern region of the grid where there is higher terrain. While the first 4 schemes produce maximum PBL depths of 1 to 1.2 km, QNSE (Fig. 43) produced PBL depths close to, and in limited areas exceeding, 1.8 km, which is at least 50% deeper than the others. Without observational data, it is not possible to say which is correct, but we do note a large increase in the PBL depth when using this scheme. Again, over higher terrain of the southeast, and in the northeast quadrant of the model domain, we find a notably deeper PBL generated using the QNSE PBL scheme. Deeper PBLs are also found within the remainder of the domain in general. So we can conclude that overall QNSE produced deeper CBLs over land at this time.

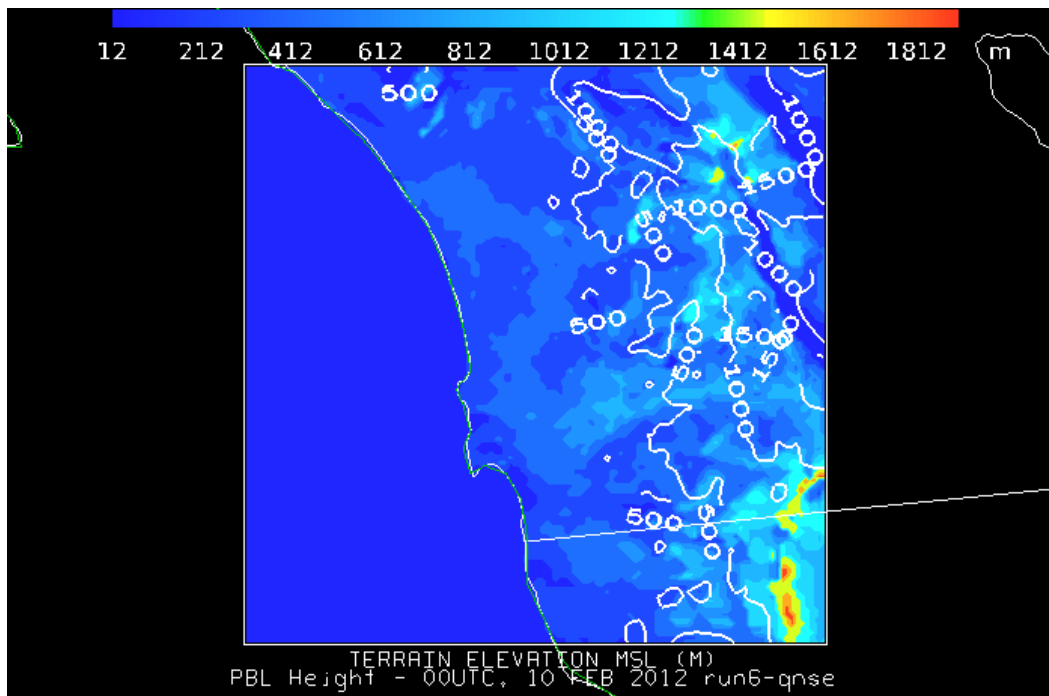


Fig. 43 PBL depth (meters) valid at 0000 UTC 10 February 2012 for the QNSE PBL option

Our final scheme tested was ACM2, and the 0000 UTC PBL depth for that case is shown in Fig. 44. The ACM2 scheme gives deeper convective PBLs than the first 5 schemes but not as deep as the QNSE scheme produced. While ACM2 produces a maximum PBL depth of around 1.8 km, the areal coverage of these deeper PBLs is much more limited than those obtained with QNSE. The deep PBL is confined to the extreme southern part of the domain, over higher terrain.

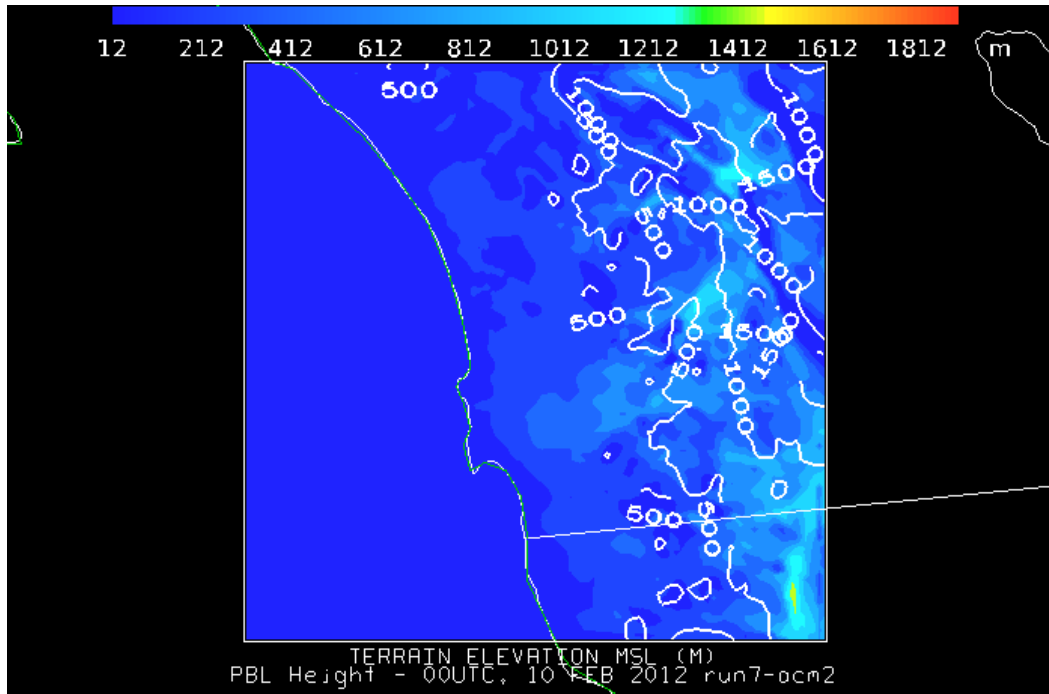


Fig. 44 PBL depth (meters) valid at 0000 UTC 10 February 2012 for the ACM2 PBL option

Our findings suggest that there is a great deal of variation in the PBL depth, and thus likely a similar level of variation in the level of the capping inversion, produced by each of the 7 schemes we analyzed. Unfortunately, we have very few validation data. To adequately test these schemes and arrive at definitive answers as to which scheme is superior, we need data from an intensive observational program that includes such tools as Lidar to examine the temporal and spatial variability and validate our model results. However, it is useful to know that such variability exists among the schemes and use this knowledge to effectively choose one particular scheme over another in decision making and battlefield forward-deployed applications, where detailed and accurate measures of PBL depth are needed.

5.5 Computation Time

The effective choice of an optimal PBL scheme to use in a forward-deployed environment with constrained computational resources and the need for rapid creation of a forecast is partly dependent on the computational cost of the scheme being considered. We compared the times required to complete the same triple nest configuration for each of the schemes. On the standalone Linux workstation we used to obtain these timings, the environment was constant during the time of integration, and no other large computational loads were being run. We ran the modeling system on a RHEL 6.9 system, a dual-core Xeon 5650 processor, and 32 GB of RAM. The results for a 24-h simulation using an 18-s time step are shown

in Fig. 45. The MYNN, SH, MYJ, BouLac, and ACM2 PBL options all ran with a wall time of about 25 h.

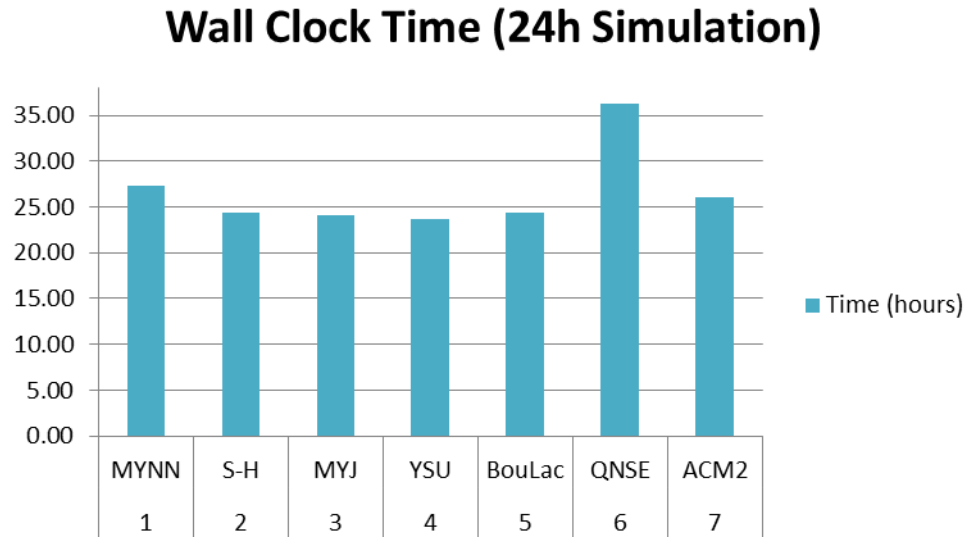


Fig. 45 Execution time (hours) for a 24-h simulation using the triple nest configuration and each of the 7 tested PBL/SL scheme combinations

The QNSE scheme took notably longer to run at approximately 35 h wall time. This could be a consideration when a strict timeline must be maintained in a forward deployment environment.

While these times are clearly not practical to be of use in nowcasting, there are a number of controllable factors that can reduce the time and make nowcasting practical. We performed experiments using an 18-s time step. Since this is quite conservative, if we employ the standard rule of thumb for stability, we chose a time step greater than or less than $6 \cdot dx$, where dx is the horizontal grid spacing that was used (the finer grids are automatically scaled down by a factor equal to the ratio of the grid spacing). We varied the time step and performed benchmarks, finally testing a 54-s time step, and the model executed in approximately 10 h. We did not evaluate the use of an adaptive time step, and that could potentially further speed up the model execution.

5.6 Analysis Summary

In this research we have analyzed the performance of 7 commonly employed WRF PBL schemes for potential use in an optimal configuration of a field-deployed application of the WRF model known as WRE-N. Analysis of these schemes showed that there is no clearly superior scheme for the meteorological variables we considered. It also showed that the best scheme is a function of meteorological

variable, time of day, and, potentially, the geographic subdomain we are interested in. No single scheme performed best in all applications. As a measure of model performance, we looked most closely at the model bias and RMSE for each of the schemes. Each of the BL schemes performed fairly similarly, with few exceptions.

Our analysis was based on a single model domain on only 1 day. However, we were able to capture the development of both the daytime CBL and the nocturnal BL as well as the transition between both. We also looked briefly at PBL depth to highlight the differences between the schemes. For the 0- to 6-h forecast period (data assimilation) during the transition between the nocturnal BL and the initial development of the daytime CBL, the BouLac scheme performed best. For the 6- to 12-h part of the simulation (nowcast), which is associated with the development of the daytime CBL, the BouLac scheme performed best. The YSU and QNSE schemes were the best performers for the extended forecast period (12–24 h) covering nighttime over the model domain. Overall, the BouLac scheme was the best performer for the forecast period.

Since we are primarily interested in nowcast applications, and since the BouLac scheme performed among the best at all times and during nighttime and daytime, we can conclude that BouLac is the preferred PBL scheme. However, this conclusion comes from a single-time configuration over 1 case day over a small region, and we only evaluated model performance based on surface data. More-extensive testing using vertical profile observations, and many more cases would be needed to definitively conclude that BouLac is preferred overall. The separation between ensemble members was small, which is reflected in the Talagrand diagrams and the standard deviation of the ensemble. We cannot confidently separate the influence of time of day from time since model initialization.

Since PBL depth is also a meteorological variable of considerable interest and value for our application, our choice would be taken with the caveat that it is pending a closer analysis of the PBL depth. Although we looked briefly at PBL depth, we need to acquire a more detailed high-temporal-resolution observational dataset such as the planned Meteorological Sensor Array (MSA) (Knapp et al. 2018). One or more movable measurement systems such as Lidar, or supplemental soundings, would be highly beneficial and provide observational data regarding the diurnal variation in the PBL depth. Such data are highly needed and invaluable in PBL model verification studies beyond using surface data.

6. Summary and Conclusions

In this series of experiments, 7 PBL/SL combinations were tested and evaluated for performance using a 9-/3-/1-km triple nest grid configuration centered over San Diego, California, under quiescent conditions in late winter. The particular domain was chosen in order to include complex and varying topography and leverage the availability of high-density surface meteorological observations, which were assimilated into the initial conditions of the model, and for detailed verification analysis. Our study was largely limited to surface parameters of temperature, dew point, and wind speed and to one day of intensive surface observations over a single model domain. The day was chosen to allow examination of the diurnal cycle of the BL in the absence of strong synoptic forcing and rapidly changing weather conditions.

For the 7 BL/SL scheme combinations we tested, we evaluated the model bias and RMSE over the innermost (1-km) model domain. For the first 6 h of the forecast period (nowcast), the mean temperature bias was -1.66 K with a standard deviation among the schemes of approximately 0.2 K. For the following 12-h extended forecast the mean/standard deviation of the bias was $0.17/0.76$ K, and for the complete forecast period it was $-0.44/0.55$ K. For dew point, the mean/standard deviation of bias for the 3 periods were, respectively, $-1.12/0.45$, $-1.06/0.19$, and $-1.10/0.15$ K. For wind speed, the respective bias statistics were $0.07/0.49$, $0.43/0.27$, and $0.31/0.34$ ms^{-1} . The smallest temperature bias for the nowcast, extended forecast, and complete forecast periods was produced by the MYJ, YSU, and MYJ schemes, respectively. For DPT it was YSU, BouLac, and BouLac. For wind, the best performers were MYNN, SH, and BouLac, respectively. In general, the spread among the tested schemes was not large; in the case of temperature and DPT, it was well under 1 K. Wind speed mean and standard deviation of bias were typically less than 0.5 ms^{-1} in light winds over the model domain.

To analyze relative performance of each scheme, we chose to focus primarily on the model RMSE since large deviations in both the positive and negative directions can offset each other in the bias statistic. The mean RMSE for surface temperature, and the corresponding standard deviation for the nowcast, extended forecast, and complete forecast periods, are $2.43/0.18$, $2.95/0.15$, and $2.77/0.13$ K, respectively. Thus, the RMSE increases with simulation time as would be expected. For DPT, these values are respectively $3.38/0.09$, $2.77/0.07$, and $2.97/0.07$ K. For wind, the mean RMSE/standard deviation pairs are $1.49/0.12$, $1.47/0.16$, and $1.47/-0.14$ ms^{-1} . For temperature, DPT, and wind, the standard deviation, or spread, in the schemes is small. From the RMSE values derived from each scheme, a ranking was developed ranging from 1 (best) to 7 (worst) and a composite developed based on

the 3 independent meteorological parameters. For the nowcast period, the best overall performer was BouLac, followed by MYJ. For the extended forecast period, the YSU, BouLac, and QNSE schemes were tied for best performance. For the overall forecast period (nowcast, extended) BouLac was best, followed by QNSE. The standard deviation of the ensemble created using the various PBL schemes was small despite the fact that the schemes we tested were devised using local, nonlocal, and hybrid approaches. For our case and grid spacing, this did not seem to make an appreciable difference in the quality of performance. While the SH scheme was developed primarily to address the scale awareness problem, we did not see a significant improvement in using this scheme compared with the others we tested, and in particular the YSU scheme, which shares common lineage. Independent statistical analysis by Smith et al. (2018) is in agreement with our determination of the lack of variance between the schemes, and attributes only 3%–4% of total model variance to the PBL schemes. That calculation involves, at the highest level, an extension of a common data analytics approach called ANOVA (for analysis of variance), which uses a parameter eta (η). Bakeman (2005) found that a generalized eta squared, η_G , was superior. Eta squared (η_G^2) was calculated using the lsr package in R (Navarro 2015), allowing us to estimate the relative contribution to model error (uncertainty) due to the BL physics packages.

Most interestingly, the variation in performance between the schemes is greater during the nighttime, suggesting that the largest differences between the formulation of the schemes is within the nighttime stable BL physics. We see this in the winds, temperature, and moisture analyses that we performed. The difference in performance of these schemes during the daytime (convective BL) is much smaller regardless of the details of the closure methodology. Further analysis should also consider examining the performance of the various BL schemes for the other 4 days, where synoptic forcing was stronger and more variable.

Curiously, the model statistics were not better with the use of the finer innermost (1-km) grid. For the grid points residing within the 1-km innermost nest, the 3-km data were slightly better for the single time analyzed but not significantly different. There may be a number of reasons for this apparent discrepancy. Decreasing grid spacing in mesoscale models to less than 10–15 km generally improves the realism of the results but does not necessarily significantly improve the objectively scored accuracy of the forecasts (Mass et al. 2002). That study further showed that position errors, timing errors, and observational data density can all contribute to these errors, and that additional approaches should be added to the current traditional verification tools. Ebert (2008) addressed this issue and attributed it to the shortcomings of using a grid-to-point verification technique, which often resulted in the conclusion that lower-resolution models scored higher in terms of continuous

error statistics than higher-resolution models despite the more realistic representation of atmospheric features provided by the latter. Errors could be caused in part by the geolocation errors documented with the WRF model; however, we have no confirmation that the spherical earth assumption has a notable effect on these simulations. Dyer et al. (2016) found, in a similar study over the same model domain, that grid spacing influences the magnitude of uncertainty within an ensemble, and that the magnitude and type of relationship varies by case.

Talagrand diagrams constructed for several hours of the simulation showed very little dispersion between the 7 members considered here relative to the spread present in the observations. Because we did not obtain an optimal flat response in our Talagrand plots, but rather a classic “U shaped” distribution characteristic of inadequate spread between the ensemble members for all but temperature, the use of only a physics-based ensemble (in this case BL/SL combinations) is, in itself, not adequate for constructing a useful ensemble to derive probabilistic information about forecast uncertainty. The Talagrand diagrams for the moisture, wind, and temperature state variables all demonstrated adequate spread to constitute a valid ensemble using only the PBL schemes as a basis for a physics-based ensemble. Therefore, it is necessary to consider more than simply the BL/SL combinations to construct an adequate ensemble from which to represent model physics uncertainty. Research looking into the relative contributions to model uncertainty due to the other physics schemes and initialization data is addressed in the DoE approach described by Smith et al. (2018).

The primary goal of this study was to examine the performance of a large number of PBL/SL combinations and potentially find the best available single preconfigured configuration that can be used in most environments and geographic regions. Although we have deliberately chosen a diverse model domain for our study, the present study does not allow the latter goal to be addressed. While we conclude that most schemes performed nearly equally well on this domain, it would be advantageous to leverage the ability to analyze subdomains using GIS tools to determine in which geographic subregion a particular scheme functions best. One possible way to address this in a field-deployable system would be to utilize a geographic database that categorizes regions. The end user would just then need to input the coordinates of the center point, and the system software would preconfigure the run using that information and the static databases such as topography and land use. Minimal user input would be required in such a scheme.

7. Further Considerations

This research explored the options presented within the WRF-ARW framework for representing PBL and SL processes and characterizing the structure of the atmosphere to produce quality deterministic forecasts within the framework of a nowcast version of the modeling system. While we have not identified a single preferred scheme for a universal configuration of the model that can be used in a variety of geographic regimes and topography, a number of other determinants has surfaced. The computational requirements of each candidate scheme differ. The most complex scheme (i.e., QNSE) requires significantly more resources to run than the simpler schemes (i.e., MYNN and YSU). This translates to longer model execution time, which impacts the practicality of using the scheme where forecast production must occur as rapidly as possible. Except at nighttime, the 7 schemes tested here did not show much variation in the measures of model performance (bias and RMSE), particularly during the daytime, so it could be argued that we should simply use the most computationally efficient code. Therefore, when designing a specific system, one might be willing to sacrifice accuracy for speed but that it depends on the specific requirements of the system.

The WRE-N modeling system is designed to produce a nowcast, or 0- to 3-h and perhaps 0- to 6-h forecast for use in a forward-deployed environment. The model output and the output from postprocessing for specific needs, like weather-impact decision aids (WIDAs), must be available to the end user rapidly enough to be useful for the decision maker and for operations. With the configuration run in these experiments, it does not appear that WRE-N would produce a timely forecast on forward-deployed hardware unless that hardware was a computer cluster. Internal testing shows execution times, as expected, are a function of the number of processors utilized. With a 12-core Xeon system (local dedicated Xeon dual-processor workstations), a 24-h run takes approximately 24 h using an 18-s time step. Execution time reduced to approximately 10 h with the suggested time step of 54 s for this configuration. Clearly, improvements in execution time are necessary for this to be a completely viable solution, but at present, without changes, the modeling system in its present form can be useful in providing nowcast support in the battlefield.

However, there are ways in which we can make the modeling system run more efficiently. First, the default time step used in this study was rather conservative. Conventional wisdom within the WRF-ARW modeling community is to use a time step $dt = 6 * dx$ via the parameter `time_step` where `parent_time_step_ratio` sets the ratio between the time step of a given nest and its parent nest. This would imply $dt = 54$ s for a 9-km outer mesh. Runs completed here used an 18-s time step (again

scaled for inner nests). The 18-s time step was likely chosen to avoid computational instability due to the complex topography of our domain and the interaction with the boundaries of our selected meshes. To test this hypothesis, an additional run was made using a 45-s time step. With fewer iterations, the model should run faster by the factor 45/18. WRF completed a 24-h simulation in 12 h and 15 min using a 45-s time step vice 29 h and 30 m using the default 18-s time step. The same simulation and configuration took 10 h and 15 min using a 54-s time step. An adaptive time step can be applied (Hutchinson 2007), but we have not yet applied this technique to this case. Second, we can speed up model execution by running on hardware configured with a solid-state drive (SSD), but if we use RAID disks this may or may not result in speedups. The SSD would minimize input–output (I/O) time writing to and reading from disk. Benchmarks were not run because the SSD drive configuration was not available at this writing. Third, the triple nest grid used in these experiments is larger than needed for typical artillery meteorology applications. Restricting the domain, particularly the outer 2 meshes, would result in further savings, as these tests were performed using a conservative grid nesting strategy to minimize the possibility of artificial features (those arising from boundary conditions) advecting inward from the outer domains. Initial testing with a 13.5-/4.5-/1.5-km domain with approximately 100×100 grid cells on each domain was able to integrate ahead 1 h in a little over 5 min on a dual CPU (8 core per CPU) desktop.

The model core itself is only one part of several pieces that have to successfully execute within given time constraints for WRE-N to be viable. Mississippi State University researchers (Dyer et al. 2015) developed a complete modeling system (ingest through postprocessing) that integrates the modeling system. Built to run on a virtual machine, the model was run on several domains including the southern California domain used here. Although the system produced a valid forecast in less than 1 h, the grid size used was far too coarse—the finest being 9 km—to address the projected needs of the US Army.

Kirby et al. (2013) describe a process by which ARL’s Battlefield Environment Division (BED) engineered a Web service that automates the entire process, from initialization data acquisition to execution of the WRF postprocessing system (WPS) via an earlier version of the RUNWPSPLUS software used in the current study (which is driven by a Perl script) to observation retrieval and quality control, and, finally, generation of high-resolution WRF grids utilizing FDDA. The WRE-N Real Time system (WREN_RT) is the latest ongoing effort to integrate the modeling process into a single framework. Part of this package includes RUNWPSPLUS, which automates a lot of the data ingest and initialization work into a single package that assimilates data, processes the model initialization, and

produces a forecast based on preselected model parameters. Postprocessing of the output data would have to follow to produce custom-tailored products for the end user, and this too can be automated to a large degree.

Modeling system testing, evaluation, and a proof of concept using real-time data, potentially within the framework of the ARL MSA (Knapp et al. 2018), would be the next step in delivering this system to the end user. A well-designed program would incorporate remote sensing, which would address the need for additional model verification and validation data, such as profiles, extending the model validation above the surface. Lidar and profilers could be used to document the diurnal evolution of the BL, a severely limiting factor in our current study.

Aside from the clear utility of using meteorological data for decision making and planning purposes, WRE-N output can be used as input to WIDAs. One such ARL application is MyWIDA (Brandt et al. 2013). Figure 46 shows an example of another application, Air Impacts Routing (AIR). In this application, thresholds are selected by the user, and a flight path solution based on the impact of specified parameters minimizes the environmental risk. To couple the AIR software to the WRF output, data conditioning must take the netCDF output and be capable of parsing it into the proper format for the calculations. Our initial testing at BED suggests this can be accomplished in a few seconds, so the postprocessing is trivial in the way of computational resources. WRF and the AIR can both be run in the field, an important consideration since transferring WRF output via currently available communication links is not practical due to the immense file size of the NetCDF output data. Processing only needed data for the intended algorithms can also improve performance by minimizing I/O, needless postprocessing, and creation of unused meteorological fields.

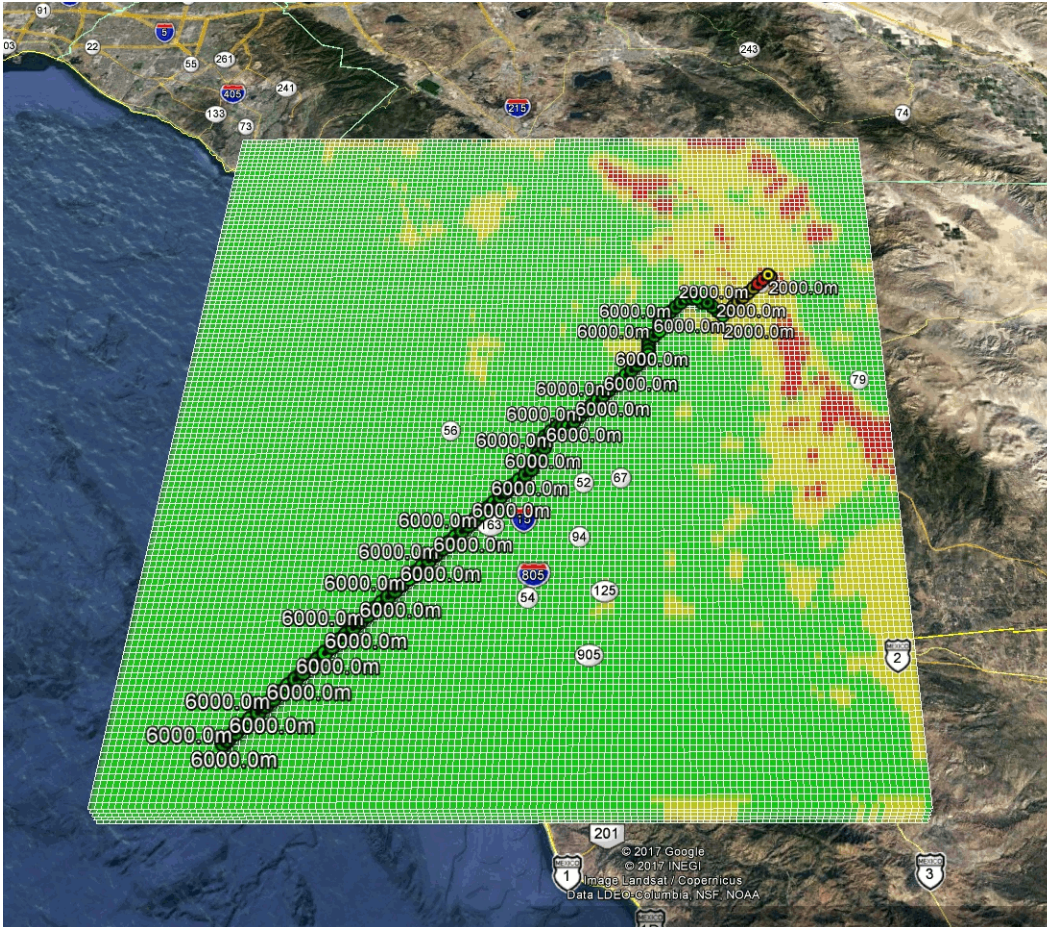


Fig. 46 Sample of WRF-ARW output after assimilation into AIR. Based on user-specified thresholds, air routing algorithms provided an optimal flight path to minimize hazards. Environmental parameters were provided by WRF-ARW output from this study. (Image courtesy of Jeff Johnson, ARL.)

While WRF embedded into a modeling system such as WRE-N is certainly possible, further improvements in model execution can be implemented to decrease the time for the components to execute and increase the lead time for end users. Clearly, there still remains a need to decrease the model bias and RMSE for the model domain. While the model bias can be mitigated by performing postprocessing adjustments in the model data, it is preferable to identify the root causes of the model bias rather than simply adjust for it. The current PBL/SL schemes produce very similar results despite the differences in the physics during the development of the daytime CBL and differ significantly with the development of a nocturnal BL. It is therefore necessary to take a closer look at the modeling system as a whole and examine the interactions between the various physical parameterizations in each category. This is the subject of ongoing research that incorporates the DoE approach, which was proposed by Smith and Penc (2015a,b, 2016) and Smith et al. (2017, 2018).

8. References

- Alpers W, Brümmer B. Atmospheric boundary layer rolls observed by the synthetic aperture radar aboard the ERS-1 satellite. *J Geophys Res.* 1994;99:12613–12621.
- Angevine WH, Jiang HT, Mauritsen T. Performance of an eddy diffusivity–mass flux scheme for shallow cumulus boundary layers. *Mon Wea Rev.* 2010;138:2895–2912.
- Angevine WM, Hare JE, Fairall CW, Wolfe DE, Hill RJ, Brewer WA, White AB. Structure and formation of the highly stable marine boundary layer over the Gulf of Maine. *J Geophys Res: Atmospheres.* 2006;111(23):2156–2202.
- Ayotte K, Sullivan P, Andren A, Doney S, Holtslag A, Large W, McWilliams J, Moeng C-H, Otte M, Tribbia J, Wyngaard J. An evaluation of neutral and convective planetary boundary layer parameterizations relative to large eddy simulations. *Boundary Layer Meteor.* 1996;79:I-3–I-175.
- Bakeman R. Recommended effect size statistics for repeated measures designs. *Behavior Research Methods.* 2005;37(3):379–384.
- Balzarini A, Angelini F, Ferrero L, Moscatelli M, Perrone MG, Pirovano G, Riva GM, Sangiorgi G, Toppetti AM, Gobbi GP, Bolzacchini E. Sensitivity analysis of PBL schemes by comparing WRF model and experimental data. *Geosci Model Dev.* 2014;7:6133–6171.
- Bougeault P, Lacarrère P. Parameterization of orography-induced turbulence in a mesobeta-scale model. *Mon Wea Rev.* 1989;117:1872–1890.
- Brandt J, Dawson L, Johnson J, Kirby S, Marlin D, Sauter D, Shirkey R, Swanson J, Szymer R, Zeng S. Second-generation weather impacts decision aid applications and web services overview. White Sands (NM): Army Research Laboratory (US); 2013 Aug. Report No.: ARL-TR-6525. p. 48.
- Bretherton C, Park S. A new moist turbulence parameterization in the community atmosphere model. *J Clim App Meteor.* 2009;22:3422–3448.
- Bright DR, Mullen SL. The sensitivity of the numerical simulation of the southwest monsoon boundary layer to the choice of PBL turbulence parameterization in MM5. *Wea Forecasting.* 2002;17:99–114.
- Bryan GH, Wyngaard JC, Fritsch JM. Resolution requirements for the simulation of deep moist convection. *Mon Wea Rev.* 2003;131:2394–2416.

- Cohen A, Cavallo S, Coniglio M, Brooks H. A review of planetary boundary layer parameterization schemes and their sensitivity in simulating southeastern US cold season severe weather environments. *Wea Forecast*. Published online; 2015 June 5. p. 591–612. <https://doi.org/10.1175/WAF-D-14-00105.1>.
- Coniglio M, Correia J Jr, Marsh P, Kong F. Verification of convection-allowing WRF model forecasts of the planetary boundary layer using sounding observations. *Wea Forecast*. Published online; 2013 Jan 6 [accessed 2018 Mar 13]. <https://journals.ametsoc.org/doi/abs/10.1175/WAF-D-12-00103.1>.
- Cuchiara C, Li X, Carvalho J, Rappengluck B. Intercomparison of planetary boundary layer parameterizations and its impacts on surface ozone concentration in the WRF/Chem model for a case study in Houston, Texas. *Atmos Env*. 2014;96:175–185.
- Dawson LP, Raby JW, Smith JA. The automation of nowcast model assessment processes. Adelphi (MD): Army Research Laboratory (US); 2016 Sep. Report No.: ARL-MR-0940.
- Deng A, Stauffer DR. On improving 4-km mesoscale model simulations. *J Appl Meteor Clim*. 2006;45:361–381.
- Deng A, Stauffer D, Gaudet B, Dudhia J, Bruyere C, Wu W, Vandenberghe F, Liu Y, Bourgeois A. Update on WRF-ARW end-to-end multi-scale FDDA system. Proceedings of 10th National Center for Atmospheric Research (NCAR) WRF Users' Workshop. 2009 June 23–26; Boulder, CO. p. 14. <http://www2.mmm.ucar.edu/wrf/users/workshops/WS2009/WorkshopPapers.php>.
- Dimitrova R, Silver Z, Fenando H, Leo L, DiSabatino S, Hocut C, Zsedrovits T. Intercomparison between different PBL options in WRF model: modification of 2 PBL schemes for stable conditions. Proceedings of the 94th Annual Meeting of the American Meteorological Society (AMS); 2014 Feb 2–6; Atlanta GA. p. 7.
- [DTC] Developmental Testbed Center. The developmental testbed center's final report on the performance of the ACM2 PBL and PX SL model in the Air Force weather agency's weather forecast system [accessed 2015 Feb 27]. https://dtcenter.org/eval/meso_mod/afwa_test/wrf_v3.6.1/AFWA2014_FINAL_REPORT.pdf.

- Dumais R, Kirby S, Flanigan R. Implementation of the WRF 4-dimensional data assimilation method of observation nudging for use as an ARL weather running estimate-nowcast. Aberdeen Proving Ground (MD): Army Research Laboratory (US); 2013 June. Report No.: ARL-TR-6485. p. 24.
- Dumais R, Passner J, Flanigan R, Sauter B, Kirby S. High resolution WRF-ARW studies at the US Army Research Laboratory for use in short-range forecast operations. Short range forecast applications: P2.4. Proceedings of the 23rd Conference on Weather Analysis and Forecasting/19th Conference on Numerical Weather Prediction; 2009 June 1–5; Omaha, NE.
- Dyer J, Zarzar C. US Army Research Laboratory (ARL)/Mississippi State University (MSU) project: atmospheric modeling and decision aids: field-based numerical weather simulations and analysis support tools period of performance: WRF on a laptop-user's guides. White Sands (NM): Battlefield Environment Division, Army Research Laboratory (US); 2015 Sep 29. p. 31.
- Dyer J, Zarzar C, Dumais R, Raby J, Smith JA. Defining the influence of horizontal grid spacing on ensemble uncertainty within a regional modeling framework. *Wea Forecast*. 2016;31:1997–2017.
- Ebert E. Fuzzy verification of high resolution gridded forecasts: a review and proposed framework. *Meteorol Appl*. 2008;15:51–64.
- Eckel FA, Glahn HR, Hamill TM, Joslyn S, Lapenta WM, Mass CF. National mesoscale probabilistic prediction: status and the way forward a white paper report from the national workshop on mesoscale probabilistic prediction; 2009 Sep [accessed 2010 May 28]. http://www.nws.noaa.gov/ost/NMPP_white_paper_28May10_with%20sig%20page.pdf.
- Etling D, Brown RA. Roll vortices in the planetary B: a review. *Boundary Layer Meteorology*. 1993;65(3):215–248.
- Fernando HJS, Pardyjak ER, Di Sabatino S, Chow FK, De Wekker SFJ, Hoch SW, Hacker J, Pace JC, Pratt T, Pu Z, et al. THE MATERHORN: unraveling the intricacies of mountain weather. *B Amer Meteor Soc*. 2015;96(2):1945–1967.
- Foley T. US Army Research Laboratory, personal communication, 2015.
- Foley T, Smith J, Raby J, Reen B, Penc R. Developing sub-domain verification methods using GIS tools. Presented at the Esri User Conference; 2015 July 21; San Diego, CA.

- Gemmill W, Katz B, Li X. Daily real-time, global sea surface temperature. High-resolution analysis: RTG_SST_HR. NOAA/NWS/NCEP/MMAB Office Note 260; 2007. p. 39. <http://polar.ncep.noaa.gov/mmab/papers/tn260/MMAB260.pdf>.
- Grenier H, Bretherton C. A moist PBL parameterization for large-scale models and its application to subtropical cloud-topped marine boundary layers. *Mon Wea Rev.* 2001;129:357–377.
- Hamill TM. Interpretation of rank histograms for verifying ensemble forecasts. *Mon Wea Rev.* 2001;129:550–560.
- HariPrasad D, Salgado R, Perdigao J, Challa V. A regional climate simulation study using WRF-ARW model over Europe and evaluation for extreme temperature weather events. *Int J Atmos Sci.* 2014;22. Article ID 704079. <http://dx.doi.org/10.1155/2014/704079>.
- Holt T, Raman S. A review and comparative evaluation of multilevel boundary layer parameterizations for first order and turbulent kinetic energy closure schemes. *Rev Geophys.* 1988;26(4):761–780.
- Holtslag AAM, Boville BA. Local versus nonlocal boundary-layer diffusion in a global climate model. *J Clim.* 1993;6:1825–1842.
- Hong S, Pan H. Nonlocal boundary layer vertical diffusion in a medium-range forecast model. *Mon Wea Rev.* 1996;124:2322–2339.
- Hong S, Noh Y, Dudhia J. A new vertical diffusion package with an explicit treatment of entrainment processes. *Mon Wea Rev.* 2006;194:2318–2341.
- Hu X-M, Klein P, Xue M. Evaluation of the updated YSU planetary boundary layer scheme within WRF for wind resource and air quality assessments. *J Geophys Res Atmospheres.* 2013;118(10):490–505.
- Hu X-M, Nielsen-Gammon J, Zhang F. Evaluation of 3 planetary boundary layer schemes in the WRF model. *J Appl Meteor Climo.* 2010;49:1831–1844.
- Huang X, Xiao Q, Barker D, Zhang X, Michalakes J, Huang W, Henderson T, Bray J, Chen Y, Ma Z, et al. Four-dimensional variational data assimilation for WRF: formulation and preliminary results. *Mon Wea Rev.* 2009;137:299–314.
- Hurrell JW, Holland MM, Gent PR. The community earth system model: a framework for collaborative research. *Bull Amer Meteor Soc.* 2013;94:1339–1360.

- Hutchinson TA. An adaptive time-step for increased model efficiency. Extended abstracts. National Center for Atmospheric Research (NCAR) 8th WRF Users' Workshop. NCAR 9.4. 2007; Boulder, CO. http://www2.mmm.ucar.edu/wrf/users/workshops/WS2007/abstracts/9-4_Hutchinson.pdf.
- Jacobs NA, Mulally DJ, Anderson AK. Correction of flux valve-based heading for improvement of aircraft wind observations. *J Atmos Oceanic Technol.* 2014;31:1733–1747.
- Janjic Z. The step-mountain eta coordinate model: further developments of the convection, vicious sublayer, and turbulence closure schemes. *Mon Wea Rev.* 1994;122:927–945.
- Jimenez P, Dudhia J. On the ability of the WRF model to reproduce the surface wind direction over complex terrain. *J Appl Meteor Clim.* 2013;52:1610–1617.
- Kaimal JC, Businger JA. Case studies of a convective plume and a dust devil. *J Appl Meteor.* 1970;9:612–620.
- Kirby SF, Reen BP, Dumais RE Jr. An automated weather research and forecasting (WRF)-based nowcasting system: software description. Aberdeen Proving Ground (MD): Army Research Laboratory (US); 2013 Oct. Report No.: ARL-SR-0283. p. 24.
- Knapp DI, Hocut CM, MacCall BT, Smith JA. The Army Research Laboratory's Atmospheric Science Center. Proceedings of the 98th Annual American Meteorological Society Meeting/19th Symposium on Meteorological Observation and Instrumentation; 2018 Jan 7–11; Austin, TX. Paper 7.2.
- Kosovic B, Curry J. A large eddy simulation study of a quasi-steady, stably stratified atmospheric boundary layer. *J Atmos Sci.* 2000;57:1052–1068.
- Liu Y, Bourgeois A, Warner T, Swerdlin S, Hacker J. Implementation of observation nudging based FDDA into WRF for supporting ATEC test operations. Presented at the 6th National Center for Atmospheric Research (NCAR) WRF/15th MM5 Users' Workshop; 2005; Boulder, CO.
- Lu C, Yuan H, Schwartz B, Benjamin S. Short-range numerical weather prediction using time-lagged ensembles. *Wea Forecast.* 2007;22:580–595.
- Mass C, Ovens D, Westrick K, Colle BA. Does increasing horizontal resolution produce more skillful forecasts? The results of two years of real-time numerical weather prediction over the Pacific Northwest. *B Amer Meteor Soc.* 2002;83:407–430.

- Mauritsen T, Svensson G, Zilitinkevich SS, Esau I, Enger L, Grisogono B. A total turbulent energy closure model for neutrally and stably stratified atmospheric boundary layers. *J Atmos Sci*. 2007;64:4113–4126.
- Mellor GL, Yamada T. A hierarchy of turbulence closure models for planetary boundary layers. *J Atmos Sci*. 1974;31:1791–1806.
- Mellor GL, Yamada T. Development of a turbulence closure model for geophysical fluid problems. *Rev Geophys Space Phys*. 1982;20:851–875.
- Monaghan A, Barlage M, Boehnert J, Phillips C, Wilhelmi O. Overlapping interests: the impact of geographic coordinate assumptions on limited-area atmospheric model simulations. *Mon Wea Rev*. 2013;141:2120–2127.
- Nakanishi M, Niino H. An improved Mellor-Yamada level 3 model: its numerical stability and application to a regional prediction of advection fog. *Boundary Layer Meteorology*. 2006;119:397–407.
- Nakanishi M, Niino H. Development of an improved turbulence closure model for the atmospheric boundary layer. *Journal of the Meteorological Society of Japan*. 2009;87(5):895–912.
- Navarro DJ. Learning statistics with R: a tutorial for psychology students and other beginners. Version 0.5. Adelaide (Australia): University of Adelaide; 2015.
- [NCAR]. National Center for Atmospheric Research. Model evaluation tools. Version 4.1 (METv4.1). User's guide 4.1. Boulder (CO): NCAR; 2013 May.
- [NOAA] Meteorological assimilation data ingest system (MADIS). College Park (MD): National Oceanic and Atmospheric Administration [accessed 2018 April 21]. <http://madis.noaa.gov>.
- Nolan D, Zhang J, Stern D. Evaluation of planetary boundary layer parameterizations in tropical cyclones by comparison of in-situ observations and high-resolution simulations of Hurricane Isabel (2003). Part I. initialization, maximum winds, and the outer-core boundary layer. *Mon Wea Rev*. 2009a;137:3651–3674.
- Nolan D, Stern D, Zhang J. Evaluation of planetary boundary layer parameterizations in tropical cyclones by comparison of in-situ observations and high-resolution simulations of Hurricane Isabel (2003). Part II. inner-core boundary layer and eyewall structure. *Mon Wea Rev*. 2009b;137:3675–3698.
- [NSIDC]. National Snow and Ice Data Center. Snow data assimilation system (SNODAS) data products at NSIDC. Version 1. Boulder (CO): NSIDC; 2004. doi:10.7265/N5TB14TC.

- Pagowski M. Some comments on PBL parameterizations in WRF. Presented at the 5th National Center for Atmospheric Research (NCAR) WRF/14th MM5 User's Workshop; 2004 June 22–25, Boulder, CO.
- Pleim J. A combined local and nonlocal closure model for the atmospheric boundary layer. Part I: model description and testing. *J Appl Meteor Climatol*. 2007a;46:1383–1395.
- Pleim J. A combined local and nonlocal closure model for the atmospheric boundary layer. Part II: application and evaluation in a mesoscale meteorological model. *J Appl Meteor Climatol*. 2007b;46:1396–1409.
- Poulos G, Blumen W, Fritts D, Lundquist J, Sun J, Burns S, Nappo C, Banta R, Newson R, Cuxart J, Terradellas E, Balsey B, Jensen M. CASES-99: a comprehensive investigation of the stable nocturnal boundary layer. *B Amer Meteor Soc*. 2002;98:555–581.
- R Core Team R: A language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing [accessed 2013]. <http://www.R-project.org/>.
- Reen BP. A brief guide to observation nudging in WRF. Boulder (CO): University Corporation for Atmospheric Research; 2016. p. 34. <http://www2.mmm.ucar.edu/wrf/users/docs/ObsNudgingGuide.pdf>.
- Reen BP, Dumais RE Jr. Assimilating tropospheric airborne meteorological data reporting (TAMDAR) observations and the relative value of other observation types. Aberdeen Proving Ground (MD): Army Research Laboratory (US); 2014 Aug. Report No.: ARL-TR-7022. p. 87.
- Reen BP, Dumais RE Jr. Assimilation of Aircraft Observations in High-Resolution Mesoscale Modeling Advances in Meteorology, Volume 2018, Article ID 8912943, 2018, p. 16
- Reen BP, Dumais RE Jr, Passner JE. Mitigating excessive drying from the use of observations in mesoscale modeling. Aberdeen Proving Ground (MD): Army Research Laboratory (US); 2014a Jan. Report No.: ARL-TR-6775. p. 33.
- Reen BP, Schmehl KJ, Young GS, Lee JA, Haupt SE, Stauffer DR. Uncertainty in contaminant concentration fields resulting from planetary boundary layer depth uncertainty. *J Appl Meteor Climatol*. 2014b;53:2610–2626.
- Salamanca F, Martilli A, Tewari M, Chen F. A study of the urban boundary layer using different urban parameterizations and high-resolution urban canopy parameters with WRF. *J Appl Meteor Climatol*. 2011;50:1107–1128.

- Schroeder AJ, Stauffer DR, Seaman NL, Deng A, Gibbs AM, Hunter GK, Young GS. Evaluation of a high-resolution, rapidly relocatable meteorological nowcasting and prediction system. *Mon Wea Rev.* 2006;134:1237–1265.
- Seaman NL, Gaudet BJ, Stauffer DR, Mahrt L, Richardson SJ, Zielonka JR, Wyngaard JC. Numerical prediction of the submesoscale flow in the nocturnal boundary layer over complex terrain. *Mon Wea Rev.* 2012;140:956–997.
- Shin H, Hong S. Intercomparison of planetary boundary-layer parametrizations in the WRF model for a single day from CASES-99. *Bound Layer Meteor.* 2011;139:261–281.
- Shin H, Hong S. Analysis of resolved and parameterized vertical transports in convective boundary layers at gray-zone resolutions. *J Atmos Sci.* 2013;70:3248–3261.
- Shin H, Hong S. Representation of the subgrid-scale turbulent transport in convective boundary layers at gray-zone resolutions. *Mon Wea Rev.* 2015;143:250–271.
- Shrivastava R, Dash S, Oza R, Sharma D. Evaluation of parameterization schemes in the WRF model for estimation of mixing height. *Int Journ Atmos Sci Vol.* 2014;9. Article No. 451578.
- Skamarock W, Klemp J, Dudhia J, Gill D, Barker D, Duda M, Huang X, Wang W, Powers J. A description of the advanced research WRF Version 3. Boulder (CO): National Center for Atmospheric Research (NCAR); 2008 June. Report No.: TN-475+STR.
- Smith JA, Penc RS. A design of experiments approach to evaluating parameterization schemes for numerical weather prediction; 2015a [accessed 2018 Mar 12]. https://www.researchgate.net/profile/Jeffrey_Smith4/publication/312166310_A_Design_of_Experiments_Approach_to_Evaluating_Parameterization_Schemes_for_Numerical_Weather_Prediction_Problem_Definition_and_Proposed_Solution_Approach/links/5873cfe008aebf17d3b09e8e/A-Design-of-Experiments-Approach-to-Evaluating-Parameterization-Schemes-for-Numerical-Weather-Prediction-Problem-Definition-and-Proposed-Solution-Approach.pdf.
- Smith JA, Penc RS. A design of experiments approach to evaluating parameterization schemes for numerical weather prediction: problem definition and proposed solution approach. *Joint Statistical Meetings Proceedings, Section on Statistics in Defense and National Security, Conference on Applied Statistics in Defense*; 2015b. p. 4183–4192.

- Smith JA, Penc R. Using design of experiments to evaluate numerical weather prediction (NWP) codes in advancing x-cutting ideas for computational climate science. Rockville (MD): Department of Energy (US); 2016.
- Smith JA, Foley TA, Raby JW, Reen B. Investigating surface bias errors in the weather research and forecasting (WRF) model using a geographic information system (GIS) by Computational and Information Sciences Directorate. Aberdeen Proving Ground (MD): Army Research Laboratory (US); 2015 Feb. Report No.: ARL-TR-7212. p. 88.
- Smith JA, Penc R, Raby JW. Statistical design of experiments in numerical weather prediction: emerging results. Presented at the 98th Annual American Meteorological Society Meeting/25th Conference on Probability and Statistics; 2018 Jan 7–11; Austin, TX. Paper No.: 6.1.
- Smith JA, Penc R, Raby JW. Is there a role for statistical design of experiments in numerical weather prediction? Presented at the 97th Annual American Meteorological Society Meeting/28th Conference on Weather Analysis and Forecasting/24th Conference on Numerical Weather Prediction; 2017 Jan 22–26; Seattle, WA. Poster No.: 616.
- Smith JA, Raby JW, Foley TA, Reen BP, Penc RS. Case study Applying GIS Tools to Verifying Forecasts over a Domain. Presented at the 96th Annual American Meteorological Society Meeting/23rd Conference on Probability and Statistics in the Atmospheric Sciences; 2016b Jan 10–14; New Orleans, LA. Paper No.: 13.3.
- Smith JA, Raby JW, Foley TA, Reen BP, Penc RS. Case study applying GIS tools to verifying forecasts over a mountainous domain. Presented at the 17th American Meteorological Society Mountain Meteorology Conference; 2016a June–1 July; Burlington, VT. Paper No.: 2.5.
- Stauffer DR. Uncertainty in environmental NWP modeling. In: Harinda J, Sermal F, editors. Handbook of environmental fluid dynamics. Volume 2. Boca Raton (FL): CRC Press; 2013. Chapter 29.
- Stauffer D, Hunter G, Deng A, Kwon Y, Hayes P, Trigg J Jr. Penn State DTRA high-resolution meteorological modeling for the Torino Winter Olympics. Presented at WRF User Workshop; 2006; Boulder, CO.
- Stensrud DJ. Parameterization schemes: keys to understanding numerical weather prediction models. New York (NY): Cambridge University Press; 2007.

- Stensrud DJ, Bao JW, Warner TT. Using initial condition and model physics perturbations in short-range ensemble simulations of mesoscale convective systems. *Mon Wea Rev.* 2000;128:2077–2107.
- Stoffler R. USAF global air land weather exploitation model. Presented at the 5th Aviation Range and Aerospace Meteorological Special Symposium/96th Annual Meeting of the American Meteorological Society. 2016 Jan 10–14; New Orleans, LA.
- Stull RB. An introduction to boundary layer meteorology. Dordrecht (The Netherlands): Kluwer Academic Publishers; 1988.
- Stull RB. Review of nonlocal mixing in turbulent atmospheres: transilient turbulence theory. *Boundary Layer Meteorology.* 1993;62:21–96.
- Sukoriansky S. Implementation of the quasi-normal scale elimination (QNSE) model of stably stratified turbulence in WRF. Beer-Sheva (Israel): Ben-Gurion University of the Negev; 2008 June. Report on WRF-DTC visit of Semion Sukoriansky; 2008 June.
- Sukoriansky S, Galperin B, Perov V. Application of a new spectral theory of stably stratified turbulence to the atmospheric boundary layer over sea ice. *Boundary Layer Meteor.* 2005;117:231–257.
- Surmeier M, Weigel J. On the optimization of the Air Force weather weapon systems next generation mesoscale numerical weather prediction model (WRF): HQ Air Force Weather Agency (AFWA). Presented at the 20th Conference on Weather Analysis and Forecasting/16th Conference on Numerical Weather Prediction/84th American Meteorological Society Meeting. 2004; Seattle, WA.
- Troen L, Mahrt L. A simple model of the atmospheric boundary layer: sensitivity to surface evaporation. *Boundary Layer Meteor.* 1986;37:129–148.
- Wang C, Jin S, Hu J, Zhang F, Feng S, Liu C. Comparing different boundary layer schemes of WRF by simulation of the low-level wind over complex terrain. *Proceedings of the International Conference on Artificial Intelligence 2*; 2011. p. 6183–6188.
- Warner TT. Numerical weather and climate prediction. Cambridge (MA): Cambridge University Press; 2011.
- Warner TT, Petersen RA, Treadon RE. A tutorial on lateral boundary conditions as a basic and potentially serious limitation to regional numerical weather prediction. *Bull Amer Meteorol Soc.* 1997;78:2599–2617.

- Wickham H. Ggplot2: elegant graphics for data analysis. Use R! New York (NY): Springer; 2009.
- Wickham H. The split-apply-combine strategy for data analysis. *Journal of Statistical Software*. 2011;40(1):1–29.
- Wickham H. Tidy data. *Journal of Statistical Software*. 2014;59(10):1–23.
- Wickham, H. Ggplot2. New York (NY): Springer Science+Business Media, LLC; 2016.
- Wickham H. Tidyverse: easily install and load the ‘Tidyverse’. R Package Version 1.2.1; 2017a [accessed 2018 Jan 18]. <https://CRAN.R-project.org/package=tidyverse>.
- Wickham H. 2017b: Ggplot2: create elegant data visualisations using the grammar of graphics. R Package Version 2.2.1; 2017b [accessed 2018 Jan 18]. <https://CRAN.R-project.org/package=ggplot2>.
- Wilczak JM, Tillman JE. The 3-dimensional structure of convection in the atmospheric boundary layer. *J Atmos Sci*. 1980;40:343–358.
- Wilczak JM, Gossard EE, Neff WD, Eberhard WL. Ground based remote sensing of the atmospheric boundary layer 25 years of progress. *Boundary Layer Meteorology*. 1996;78:321–49.
- Wilks DS. On the reliability of the rank histogram. *Mon Wea Rev*. 2011;139:311–316.
- Wyngaard JC. Toward numerical modeling in the “terra incognita. *J Atmos Sci*. 2004;61:1816–1826.
- Wyngaard JC, Brost RA, Top-down and bottom-up diffusion of a scalar in the convective boundary layer. *Journal of the Atmospheric Sciences*. 1984;41:102–112.
- Zupanski M, Navon IM, Zupanski D. The maximum likelihood ensemble filter as a non-differentiable minimization algorithm. *QJR Meteorol Soc*. 2008;134:1039–1050.

List of Symbols, Abbreviations, and Acronyms

σ	standard deviation
1-D	1-dimensional
3-D	3-dimensional
4-D	4-dimensional
ABL	atmospheric boundary layer
ACM2	Asymmetric Convection Model 2
AFWA	Air Force Weather Agency (now 557th Weather Wing)
AGL	above ground level
AIR	Air Impacts Routing
ARL	US Army Research Laboratory
BED	Battlefield Environment Directorate
BL	boundary layer
BouLac	Bougeault-Lacarrère
CASES-99	The Cooperative Atmospheric Surface Exchange Study 1999
CBL	convective boundary layer
COMP	composite
CU	cumulus
DoE	Design of Experiments
DPT	dew point temperature
DTC	Developmental Testbed Center
FA	free atmosphere/air
FDDA	Four Dimensional Data Assimilation
GALWEM	Global Air Land Weather Exploitation Model
GB	gigabyte
GFS	Global Forecast System

GIS	Geographic Information System
GRIB	gridded binary
HRRR	High Resolution Rapid Refresh (Model)
I/O	input-output
K	Kelvin
KNKX	Miramar Marine Corps Air Station (San Diego) identifier
KNZY	North Island Naval Air Station (San Diego) identifier
KSAN	San Diego (CA) identifier
KSFO	San Francisco (CA) identifier
LES	large eddy simulation
Lidar	Light Detection and Ranging
LSM	Land Surface Model
LW	longwave (infrared) radiation
MADIS	Meteorological Assimilation Data Ingest System
MATERHORN	Mountainous Terrain Atmospheric Modeling and Observation
MET	Model Evaluation Tools
ML	mixed layer
MM4	Mesoscale Model Version 4 (Penn State/NCAR)
MM5	Mesoscale Model Version 5 (Penn State/NCAR)
MO	Monin-Obukhov
MP	microphysics
MRF	Medium Range Forecast
MSA	Meteorological Sensor Array
MYJ	Mellor-Yamada-Janjic
MYNN	Mellor-Yamada Nakanishi Niino
MyWIDA	My Weather Impacts Decision Aid
NAD83	North American Datum of 1983

NCAR	National Center for Atmospheric Research
NCEP	National Center for Environmental Prediction
netCDF	Network Common Data Form
NOAA	National Oceanic and Atmospheric Administration
NOAH	NCEP, Oregon State, Air Force, Hydrology Laboratory
NSSL	National Severe Storms Laboratory
NWP	Numerical Weather Prediction
Perl	Practical Extraction Report Language
PBL	planetary boundary layer
PST	Pacific Standard Time
PX	Pleim-Xiu
QNSE	Quasi-normal Scale Elimination
RAID	redundant array of individual disks
RAM	random access memory
RH	relative humidity
RMSE	root mean squared error
RRTM	Rapid Radiative Transfer Model
SBL	stable boundary layer
SCU	shallow cumulus
SH	Shin-Hong
SL	surface layer
SSD	solid-state drive
SW	shortwave (or solar) radiation
T	temperature
TAMDAR	Tropospheric Airborne Meteorological Data Reporting
TEMF	Total Energy Mass Flux
TKE	turbulent kinetic energy

u	east–west component of wind
UPP	Unified Postprocessor
UTC	coordinated universal time
UW	University of Washington
v	north–south component of wind
WGS84	World Geodetic System 1984
WIDA	weather impact decision aid
WIND	speed variable
WPS	Weather Research and Forecast Model (WRF) postprocessing system
WRE-N	Weather Running Estimate–Nowcast Model
WRF	Weather Research and Forecasting Model
WRF-ARW	Advanced Research version of the Weather Research and Forecasting Model
WREN-RT	Weather Running Estimate–Nowcast Model, Real Time
YSU	Yonsei University

1 DEFENSE TECHNICAL
(PDF) INFORMATION CTR
DTIC OCA

2 DIR ARL
(PDF) IMAL HRA
RECORDS MGMT
RDRL DCL
TECH LIB

1 GOVT PRINTG OFC
(PDF) A MALHOTRA

1 ARL
(PDF) RDRL CIE M
R PENC

INTENTIONALLY LEFT BLANK.